



(12) 发明专利申请

(10) 申请公布号 CN 112867798 A

(43) 申请公布日 2021.05.28

(21) 申请号 201980069706.4

(22) 申请日 2019.08.30

(30) 优先权数据

62/725,225 2018.08.30 US

(85) PCT国际申请进入国家阶段日

2021.04.22

(86) PCT国际申请的申请数据

PCT/US2019/049157 2019.08.30

(87) PCT国际申请的公布数据

WO2020/047472 EN 2020.03.05

(71) 申请人 国家儿童医院研究所

地址 美国俄亥俄州

(72) 发明人 斯科特·艾伦·卢伊勒

(74) 专利代理机构 广州文冠倪律知识产权代理  
事务所(普通合伙) 44348

代理人 何锦标

(51) Int.Cl.

C12P 19/04 (2006.01)

C12P 21/02 (2006.01)

A61K 38/16 (2006.01)

A61K 38/17 (2006.01)

A61K 38/45 (2006.01)

C07K 14/075 (2006.01)

C07K 14/47 (2006.01)

C12N 5/071 (2006.01)

C12N 15/09 (2006.01)

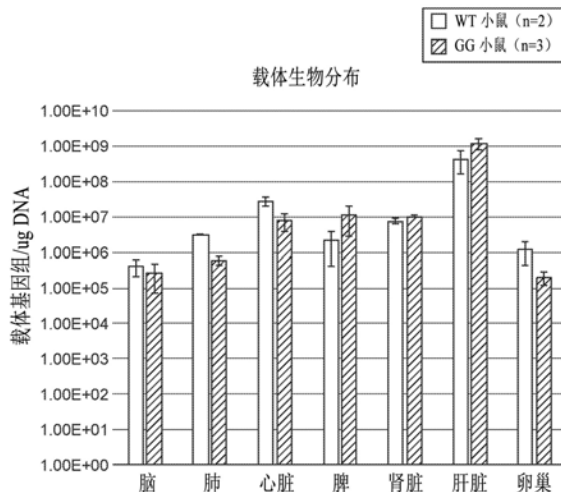
权利要求书3页 说明书40页 附图7页

(54) 发明名称

用于治疗半乳糖血症的基因疗法

(57) 摘要

本文提供了编码GALT蛋白的重组AAV载体。  
在本公开中还提供了用于治疗半乳糖血症的组  
合物和方法。



1. 一种分离的多核苷酸,其包含与SEQ ID NO:1-3中的任何一个至少85%相同的序列或其每一个的等效物,其中任选地已经从野生型序列中被修饰的核苷酸中的至少一个或多个没有从其相应的野生型序列中被修饰。

2. 一种分离的多核苷酸,其编码SEQ ID NO:4的氨基酸或与SEQ ID NO:4的氨基酸序列至少85%相同的氨基酸或其等效物。

3. 一种多核苷酸,其包含如权利要求1或2所述的多核苷酸的互补序列。

4. 一种多核苷酸,其包含SEQ ID NO:1的多核苷酸或其等效物,并且任选地其中已经从野生型序列中被修饰的核苷酸中的至少一个或多个没有从SEQ ID NO:1中被修饰。

5. 一种载体,其包含如权利要求2至4中任一项所述的多核苷酸。

6. 根据权利要求5所述的载体,其进一步包含启动子和/或增强子元件。

7. 根据权利要求1至4中任一项所述的分离的多核苷酸,或根据权利要求5或6所述的载体,其进一步包含标记。

8. 根据权利要求5所述的载体,其中所述载体是AAV载体。

9. 根据权利要求8所述的载体,其进一步包含编码衣壳蛋白的多核苷酸。

10. 根据权利要求9所述的载体,其中所述衣壳蛋白是野生型衣壳蛋白或突变的衣壳蛋白。

11. 一种分离的细胞,其包含如权利要求1至4中任一项所述的分离的多核苷酸或如权利要求5至10中任一项所述的载体。

12. 一种重组AAV载体,其包含编码半乳糖-1-磷酸尿苷酰转移酶("GALT")的多核苷酸序列。

13. 根据权利要求12所述的重组AAV载体,其中编码GALT的多核苷酸序列包含与SEQ ID NO:1至少85%相同的核苷酸序列,并且任选地其中已经从野生型序列中被修饰的核苷酸中的至少一个或多个没有从SEQ ID NO:1中被修饰。

14. 根据权利要求12所述的重组AAV载体,其中编码GALT的多核苷酸序列包含如SEQ ID NO:1所示的核苷酸序列。

15. 根据权利要求12至14中任一项所述的重组AAV载体,其中所述载体为血清型AAV1、AAV2、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、AAV13、AAV PHP.B或AAV rh74。

16. 根据权利要求12至14中任一项所述的重组AAV载体,其中所述载体是AAV9、rh74、突变的rh74或AAV PHP.B。

17. 根据权利要求12至16中任一项所述的重组AAV载体,其中所述载体可操作地连接至启动子、组织特异性控制元件或组成型启动子。

18. 根据权利要求17所述的重组AAV载体,其中所述组成型启动子包含Rous肉瘤病毒(RSR)LTR启动子(任选地具有RSV增强子)、巨细胞病毒(CMV)启动子、SV40启动子、二氢叶酸还原酶启动子、 $\beta$ -肌动蛋白启动子、磷酸甘油激酶(PGK)启动子、U6启动子或EF1启动子。

19. 根据权利要求18所述的重组AAV载体,其中所述 $\beta$ -肌动蛋白启动子是鸡 $\beta$ -肌动蛋白("CBA")启动子。

20. 根据权利要求12所述的重组AAV载体,其包含与SEQ ID NO:2至少95%相同的核苷酸序列,并且任选地其中已经从野生型序列中被修饰的核苷酸中的至少一个或多个没有从

SEQ ID NO:2中被修饰。

21. 根据权利要求12所述的重组AAV载体,其包含如SEQ ID NO:2所示的核苷酸序列。

22. 根据权利要求12所述的重组AAV载体,其包含与SEQ ID NO:3至少85%相同的核苷酸序列,并且任选地其中已经从野生型序列中被修饰的核苷酸中的至少一个或多个没有从SEQ ID NO:3中被修饰。

23. 根据权利要求12所述的重组AAV载体,其包含如SEQ ID NO:3所示的核苷酸序列。

24. 根据权利要求12至17中任一项所述的重组AAV载体,其中编码GALT酶的多核苷酸序列编码SEQ ID NO:4的氨基酸序列。

25. 根据权利要求12至24中任一项所述的重组AAV载体,其进一步包含可检测或纯化标记。

26. 一种病毒颗粒,其包含如权利要求12至24中任一项所述的AAV载体和一种或多种衣壳蛋白。

27. 根据权利要求26所述的病毒颗粒,其中所述衣壳蛋白包含至少一种野生型衣壳蛋白。

28. 根据权利要求26所述的病毒颗粒,其中所述衣壳蛋白包含至少一种非天然存在的或突变的蛋白。

29. 根据权利要求26所述的病毒颗粒,其中所述衣壳蛋白包含至少一种野生型衣壳蛋白和至少一种突变的衣壳蛋白。

30. 根据权利要求26所述的病毒颗粒,其中所述衣壳蛋白包含含有一种或多种选自以下的修饰的经修饰的AAVrh74 VP1衣壳蛋白:在AAVrh74的VP1的502位氨基酸处异亮氨酸取代天冬酰胺、以及任选地在氨基酸505处色氨酸被取代为精氨酸、以及在AAVrh74的VP1的591位氨基酸处插入肽YIG。

31. 一种多核苷酸,其编码包含一种或多种选自以下的修饰的经修饰的AAVrh74 VP1衣壳蛋白:在AAVrh74的VP1的502位氨基酸处异亮氨酸取代天冬酰胺、在氨基酸505处色氨酸被取代为精氨酸、以及在AAVrh74的VP1的591位氨基酸处插入肽YIG、或其每一个的等效物;或者至少85%相同的多核苷酸,条件是一个或多个修饰与所述经修饰的AAVrh74 VP1衣壳蛋白中的氨基酸相同。

32. 一种经修饰的AAVrh74 VP1衣壳蛋白,其包含一种或多种选自以下的修饰:在AAVrh74的VP1的502位氨基酸处异亮氨酸取代天冬酰胺、在氨基酸505处色氨酸被取代为精氨酸、以及在AAVrh74的VP1的591位氨基酸处插入肽YIG、或其每一个的等效物;或至少85%相同的多核苷酸,条件是一个或多个修饰与所述经修饰的AAVrh74 VP1衣壳蛋白中的氨基酸相同。

33. 一种如权利要求31所述的多核苷酸的互补序列。

34. 一种组合物,其包含如权利要求1至14、31或33中任一项所述的多核苷酸、如权利要求32所述的衣壳蛋白、和/或如权利要求15至19中任一项所述的病毒颗粒和载体,所述载体任选地是药学上可接受的载体。

35. 一种用于将功能性GALT酶引入细胞的方法,其包含使所述细胞与如权利要求1至13、31或33中任一项所述的多核苷酸、如权利要求32所述的衣壳蛋白、如权利要求14所述的组合物、或如权利要求15至19中任一项所述的病毒颗粒接触。

36. 根据权利要求35所述的方法,其中所述接触是离体的或体内的。

37. 一种用于在有此需要的对象中治疗半乳糖血症的方法,其包含向所述对象施用治疗有效量的如权利要求1至13中任一项所述的多核苷酸、或如权利要求14所述的组合物、或如权利要求15至19中任一项所述的病毒颗粒。

38. 一种在患有半乳糖血症的对象中增加半乳糖代谢的方法,其包含向所述对象施用如权利要求1至13中任一项所述的多核苷酸、或如权利要求14所述的组合物、或如权利要求15至19中任一项所述的病毒颗粒。

39. 一种在患有半乳糖血症的对象中减轻疾病状况的方法,其包含向所述对象施用如权利要求1至13中任一项所述的多核苷酸、或如权利要求14所述的组合物、或如权利要求15至19中任一项所述的病毒颗粒,其中所述疾病状况选自黄疸、肝脾肿大、肝细胞功能不全、低血糖、肾小管功能障碍、肌张力减退、败血症、白内障、共济失调、震颤、骨密度降低或原发性卵巢功能不全。

40. 根据权利要求37至39中任一项所述的方法,其中所述半乳糖血症是1型半乳糖血症。

41. 根据权利要求37至40中任一项所述的方法,其中所述对象是哺乳动物。

42. 根据权利要求41所述的方法,其中所述对象是人类。

43. 根据权利要求42所述的方法,其中所述对象是人类新生儿。

44. 根据权利要求43所述的方法,其中通过针对功能性GALT基因的缺乏表达的基因测试,已经选择了所述人类新生儿进行施用。

45. 根据权利要求37至44中任一项所述的方法,其中所述施用是通过肌肉内注射、局部递送至肝脏、或静脉内注射。

46. 根据权利要求37至44中任一项所述的方法,其中所述施用包含全身性施用。

47. 根据权利要求37至44所述的方法,其中所述施用包含局部递送至肝脏。

48. 一种试剂盒,其包含如权利要求1至13、31或33中任一项所述的多核苷酸、如权利要求32所述的衣壳、如权利要求14所述的组合物、或如权利要求15至19中任一项所述的病毒颗粒中的一种或多种,以及使用说明。

49. 一种宿主细胞,其包含如权利要求1至13中任一项所述的重组AAV载体。

50. 根据权利要求49所述的宿主细胞,其中所述细胞是原核或真核细胞。

51. 根据权利要求50所述的宿主细胞,其中所述细胞是包装细胞。

52. 一种AAV包装系统,其包含如权利要求1至13、31或33中任一项所述的多核苷酸和辅助细胞系。

## 用于治疗半乳糖血症的基因疗法

### [0001] 相关申请的交叉引用

本申请根据35 U.S.C. § 119(e) 要求于2018年8月30日提交的美国临时序列号62/725,225的优先权,其全部内容通过引用合并于此。

### [0002] 背景

本公开总体上涉及基因疗法领域,尤其涉及半乳糖血症的治疗。

[0003] 除了饮食限制含半乳糖的食物外,半乳糖血症患者目前尚无其他的治疗选择。即使严格遵守无半乳糖饮食,由于食物中含有少量半乳糖和内源性产生的半乳糖,许多患者仍患有长期的精神和身体缺陷。因此,本领域中需要有效的治疗。本公开满足了这种需求并且还提供了相关的优点。

### [0004] 概述

通过有效地将GALT基因传递至包括脑在内的细胞,来提供功能性GALT蛋白的持续表达并减轻疾病症状,所公开的基因疗法为半乳糖血症患者提供了持续的长期治疗选择。

[0005] 为了实现该疗法,本文提供了一种多核苷酸或重组腺相关病毒(“AAV”)载体,其包含编码半乳糖-1-磷酸尿苷酰转移酶(“GALT”)的多核苷酸序列,或基本上由其组成,或进一步由其组成。在一方面,编码GALT的多核苷酸序列包含与SEQ ID NO:1至少85%相同的核苷酸序列,或基本上由其组成,或进一步由其组成。在一方面,序列与SEQ ID NO:1至少85%相同,条件是已经从野生型序列中被修饰的核苷酸中的至少一个或多个没有从SEQ ID NO:1中被修饰(参见图4)。在另一方面,编码GALT的多核苷酸序列包含SEQ ID NO:1所示的核苷酸序列,或基本上由其组成,或进一步由其组成。在另一个实施方案中,多核苷酸序列编码SEQ ID NO:4的氨基酸序列或其等效物。

[0006] 在一方面,载体是自互补载体或单链DNA(ssDNA)载体。

[0007] 在一方面,重组腺相关病毒(“AAV”)包含AAV(例如侧翼为两个反向末端重复序列(ITR)的scAAV或ssAAV),或基本上由其组成,或进一步由其组成。这些ITR在序列的末端形成发夹来作为引物,从而在随后的感染步骤开始之前开始第二链的合成。第二链合成被认为是有效感染的几个阻碍之一。scAAV的其他优点包括体外和体内转基因表达的增加和延长。因此,在一方面,AAV还包含两个ITR。

[0008] 用于产生载体的重组AAV骨架的非限制性实施例包括选自以下血清型的AAV载体: AAV1、AAV2、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、AAV13、AAV PHP.B或AAV rh74。在另一方面,载体骨架是AAV9血清型、rh74血清型或经修饰的AAVrh74血清型。本发明还提供一种多核苷酸,其编码包含一种或多种选自以下的修饰的经修饰的AAVrh74 VP1衣壳蛋白:在AAVrh74的VP1的502位氨基酸处异亮氨酸取代天冬酰胺、以及任选地在氨基酸505处色氨酸被取代为精氨酸、以及在AAVrh74的VP1的591位氨基酸处插入肽YIG或YIGSR、或其每一个的等效物;或者至少85%相同的多核苷酸,条件是一个或多个、或两个或更多个、或三个或更多个、或四个或更多个修饰与所述经修饰的AAVrh74 VP1衣壳蛋白中的氨基酸相同;以及这些多核苷酸的互补序列。本发明还提供一种经修饰的AAVrh74 VP1衣壳蛋白,其包含一种或多种选自以下的修饰:在AAVrh74的VP1的502位氨基酸处异亮氨酸取代

天冬酰胺、在氨基酸505处色氨酸被取代为精氨酸、以及在AAVrh74的VP1的591位氨基酸处插入肽YIG、或其每一个的等效物；或至少85%相同的多核苷酸，条件是一个或多个、或两个或更多个、或三个或更多个、或四个修饰与所述经修饰的AAVrh74 VP1衣壳蛋白中的氨基酸相同。

[0009] 编码GALT的多核苷酸序列任选地可操作地连接至启动子、组织特异性控制元件或组成型启动子。组成型启动子的非限制性实施例包括，例如，Rous肉瘤病毒(RSV)LTR启动子(任选地具有RSV增强子)、巨细胞病毒(CMV)启动子、SV40启动子、二氢叶酸还原酶启动子、 $\beta$ -肌动蛋白启动子、磷酸甘油激酶(PGK)启动子或EF1启动子。在一个特定方面， $\beta$ -肌动蛋白启动子是鸡 $\beta$ -肌动蛋白(“CBA”)启动子。在另一个实施方案中，启动子选自CMV启动子、EF1a启动子、SV40启动子、PGK1(人或小鼠)启动子、P5启动子、Ubc启动子、人 $\beta$ 肌动蛋白启动子、CAG启动子、TRE启动子、UAS启动子、Ac5启动子、多角体蛋白启动子、CaMKIIa启动子、Gal11启动子、TEF1、GDS启动子、ADH1启动子、CaMV35S启动子、Ubi启动子、H1启动子、U6启动子或 $\alpha$ -1-抗胰蛋白酶启动子。

[0010] 在另一方面，多核苷酸或重组AAV进一步包含编码增强子元件的多核苷酸。非限制性实施例包括CMV增强子、WPRE和RSV增强子。

[0011] 在一个特定的方面，重组多核苷酸或AAV载体包含与SEQ ID NO:2或SEQ ID NO:3至少85%相同的核苷酸序列，或基本上由其组成，或进一步由其组成。在一方面，序列与SEQ ID NO:2或3至少85%相同，条件是已经从野生型GALT中被修饰的核苷酸中的至少一个或多个(参见图4)没有从SEQ ID NO:2或3中被修饰。在另一方面，重组AAV载体包含SEQ ID NO:2或SEQ ID NO:3所示的核苷酸序列，或基本上由其组成，或进一步由其组成。

[0012] 如本文所述的重组多核苷酸或AAV载体可以进一步包含可检测的或纯化标记。多核苷酸或载体可以作为包含多核苷酸或媒介和载体(例如防腐剂或药学上可接受的载体)的组合物而被包含。

[0013] 本发明还提供了包含如本文所述的多核苷酸和/或AAV载体的细胞或病毒颗粒。细胞可以是原核的或真核的，并且可以包括包装细胞系。细胞和病毒颗粒可以作为包含媒介和载体(例如防腐剂或药学上可接受的载体)的组合物而被包含。

[0014] 多核苷酸和载体对治疗或其他方法有用，并被包含在野生型或经修饰的衣壳蛋白中。因此，本文提供包含本文公开的重组多核苷酸和/或AAV载体的AAV病毒颗粒，以及用于本文公开的方法的含有所述多核苷酸和/或病毒颗粒的组合物。本发明还提供了制备载体、多核苷酸和包含它们的病毒颗粒的方法。在一方面，本发明提供了一种用于将功能性GALT酶引入细胞的方法，其包含将细胞与如本文所述的重组多核苷酸和/或AAV载体或包含其的衣壳或病毒颗粒或组合物接触。接触可以是离体的(*ex vivo*)或体内的。

[0015] 本发明还提供了一种用于在有此需要的对象中治疗半乳糖血症的方法，该方法包含向该对象施用有效量的如本文所述的重组多核苷酸和/或AAV载体、包含重组多核苷酸和/或AAV载体的病毒颗粒、或组合物，或基本上由其组成，或进一步由其组成。

[0016] 本发明还提供了一种在可能患有半乳糖血症的对象中增加半乳糖代谢的方法，在一方面，该方法包含向该对象施用如本文所述的重组多核苷酸和/或AAV载体、病毒颗粒或组合物，或基本上由其组成，或进一步由其组成。在一方面，所递送的量是由治疗专业人员确定的有效量。

[0017] 本发明进一步提供了一种在患有半乳糖血症的对象中减轻与半乳糖血症有关的疾病状况或症状的方法,该方法包含向该对象施用如本文所述的重组多核苷酸和/或AAV载体、病毒颗粒或组合物,或基本上由其组成,或进一步由其组成,其中所述疾病状况包含黄疸、肝脾肿大、肝细胞功能不全、低血糖、肾小管功能障碍、肌张力减退、败血症、白内障、共济失调、震颤、骨密度降低、成年女性的生育力、GalT缺陷者的运动功能受损和生长受限、或者原发性卵巢功能不全中的一种或多种。在一方面,该方法进一步包含将对象鉴定为患有半乳糖血症,然后治疗通过该方法鉴定的对象。在一方面,所递送的量是由治疗专业人员确定的有效量。

[0018] 在这些方法的每一种中,待治疗的半乳糖血症是1型半乳糖血症、2型半乳糖血症或3型半乳糖血症。在一个特定的方面,半乳糖血症是1型半乳糖血症。

[0019] 通过肌肉内注射或静脉内注射施用重组多核苷酸和/或AAV载体、病毒颗粒或组合物。或者,全身性施用重组多核苷酸和/或AAV载体/多核苷酸和/或病毒颗粒或组合物。此外,通过注射、输注或植入肠胃外施用重组多核苷酸和/或AAV载体、病毒颗粒或组合物。在另一方面,通过肌肉内注射或静脉内注射,然后随后全身性施用多核苷酸、重组AAV载体、病毒颗粒或组合物。

[0020] 本公开还提供了一种试剂盒,该试剂盒包含如本文所述的重组多核苷酸和/或AAV载体、病毒颗粒和/或组合物。试剂盒可以任选地包含用于制备和使用多核苷酸、载体、病毒颗粒或组合物的说明。

## 附图说明

[0021] 图1A-1C示出了三个示例性载体的图:pAAV-CB-GALT-WPRE-kan(图1A),pAAV-CB-hGALT(图1B)和pAAV-CB-hGALT-WPREv2(图1C)。

[0022] 图2示出了pAAV-CB-GALT-WPRE-kan载体(图1A)的生物分布。

[0023] 图3示出了使用pAAV-CB-GALT-WPRE-Kan载体(图1A)的载体表达的人GALT蛋白的表达。

[0024] 图4是与相应的野生型序列相比,编码如本文所述的GALT的密码子优化的重组多核苷酸的序列比对。

[0025] 图5示出了AAV9载体的生物分布。向C57Bl/6J(WT)或GalT<sup>-/-</sup>(CG)雌性小鼠静脉内注射1E+12的AAV9/lucEYFP报告病毒的病毒基因组(Vg),并在1周后收获组织并通过qPCR分析载体基因组拷贝。

[0026] 图6示出了蛋白质表达。将质粒pAAV-CB-GALT-WPRE或pAAV-CB-lucEYFP报告质粒转染至HEK293细胞中,并两天后通过蛋白质印迹分析测定GALT表达。GALT转染的细胞裂解液(30或3 μg)、LucEYFP细胞裂解液(138 μg)、HepG2内源性GALT蛋白对照(135 μg)、纯化的细菌表达的GALT蛋白标准品(由于纯化标签而具有稍高的分子量)。

[0027] 详细说明

在下文中将更全面地描述根据本公开的实施方案。然而,本公开的各方面可以以不同的形式实施,并且不应被解释为限于本文阐述的实施方案。而是,提供这些实施方案使得本公开将是透彻和完整的,并将向本领域技术人员充分传达本发明的范围。在本文的描述中使用的术语仅出于描述特定实施方案的目的,而无意进行限制。

[0028] 除非另有定义,否则本文中使用的所有术语(包括技术和科学术语)具有与本发明所属领域的普通技术人员通常所理解的含义。将进一步理解的是,术语(例如在常用词典中定义的那些)应该被解释为具有与其在本申请和相关技术的上下文中的含义一致的含义,并且不应该以理想化或过于正式的意义来解释,除非在本文中明确地定义。尽管以下没有明确定义,但应该根据其常用含义来解释这些术语。

[0029] 在本文的描述中使用的术语仅出于描述特定实施方案的目的,而无意限制本发明。本文提及的所有出版物、专利申请、专利和其他参考文献通过引用整体并入本文。

[0030] 除非另有说明,否则本技术的实践将采用组织培养、免疫学、分子生物学、微生物学、细胞生物学和重组DNA的常规技术,它们在本领域技术范围内。

[0031] 除非上下文另外指出,否则具体意图是可以以任何组合使用本文描述的本发明的各种特征。此外,本公开还预期在一些实施方案中,可以排除或省略本文阐述的任何特征或特征的组合。为了说明这一点,如果说明书指出复合物包含组分A、B和C,则具体意图是A、B或C中的任何一个或其组合可以以单数形式或以任何组合形式被省略和放弃。

[0032] 除非另外明确指出,否则所有指定的实施方案、特征和术语意图包括所列举的实施方案、特征或术语及其生物学等效物。

[0033] 所有数字值(包括范围),例如pH、温度、时间、浓度和分子量,都是适当地变化(+)或(-)1.0或0.1的增值的近似值,或者变化 $\pm 15\%$ 、或者 $10\%$ 、或者 $5\%$ 、或者 $2\%$ 。应当理解的是,尽管并非总是明确地说明,但所有数字值前面都有术语“约”。还应当理解的是,尽管并非总是明确说明,但本文所述的试剂仅是示例性的,并且其等效物在本领域中是已知的。

[0034] 在整个本公开中,通过识别引用或通过阿拉伯数字引用了各种出版物、专利和公开的专利说明书。可以在权利要求书的前面找到对由阿拉伯数字标识的出版物的完整引用。这些出版物、专利和公开的专利说明书的公开内容通过引用整体并入本公开中,以更全面地描述本发明所涉及的技术水平。

[0035] 定义

除非另有说明,否则本技术的实践将采用有机化学、药理学、免疫学、分子生物学、微生物学、细胞生物学和重组DNA的常规技术,这是在本领域的技术范围内。参见,例如, Sambrook, Fritsch and Maniatis, *Molecular Cloning: A Laboratory Manual*, 2nd edition (1989); *Current Protocols In Molecular Biology* (F. M. Ausubel, et al. eds., (1987)); the series *Methods in Enzymology* (Academic Press, Inc.): *PCR 2: A Practical Approach* (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) *Antibodies, a Laboratory Manual*, 以及 *Animal Cell Culture* (R.I. Freshney, ed. (1987))。

[0036] 除非上下文另有明确说明,否则如本发明的描述和所附权利要求书中所用,单数形式“一个”、“一种”和“该”旨在也包括复数形式。

[0037] 如本文使用的,术语“包含”旨在意为组合物和方法包括所述的元素但不排除其他元素。如本文中所使用的,“基本上由……组成”(和语法变体)的过渡短语应被解释为包含所列举的材料或步骤以及那些不实质地影响所列举的实施方案的基本和新颖特征的材料或步骤。因此,本文所使用的术语“基本上由……组成”不应当被解释为等同于“包含”。“由……组成”应当意为排除多于微量元素的其他成分和用于施用本文公开的组合物的实

质性方法步骤。这些过渡术语中的每一个所定义的方面都在本公开的范围內。

[0038] 如本文使用的,当涉及可测量值(例如量或浓度等)时,术语“约”意在包含指定量的20%、10%、5%、1%、0.5%、或者甚至0.1%的变化。

[0039] 当用于描述本文公开的任何组分、范围、剂型等的选择时,术语或“可接受的”、“有效的”或“足够的”意指所述组分、范围、剂型等适合于所公开的目的。

[0040] 同样如本文使用的,“和/或”是指并且包含一个或多个相关联的所列项目的任何和所有可能的组合,以及当以替代的(“或”)方式解释时缺乏组合。

[0041] 如本文使用的,术语“腺相关病毒”或“AAV”是指与此名称相关并属于细小病毒科依赖细小病毒属的病毒类别的成员。已知该病毒的多种血清型适合用于基因递送;所有已知的血清型都可以感染来自各种组织类型的细胞。至少11种顺序编号的AAV血清型是本领域已知的。用于本文公开的方法的非限制性示例性血清型包括11种血清型中的任一种,例如AAV2、AAV8、AAV9、或变体血清型,例如AAV-DJ和AAV PHP.B。AAV颗粒包含三种主要的病毒蛋白:VP1、VP2和VP3。在一个实施方案中,AAV是指血清型AAV1、AAV2、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、AAV13、AAV PHP.B或AAV rh74。

[0042] 术语“半乳糖血症”是指影响对象代谢半乳糖的能力的遗传疾病。食物(例如乳制品)中的乳糖被乳糖酶分解为葡萄糖和半乳糖。在半乳糖血症的个体中,半乳糖进一步代谢所需的酶(半乳糖激酶和半乳糖-1-磷酸尿苷酰转移酶)被严重减少或完全丧失,导致在经典型半乳糖血症的情况下,各种组织中半乳糖或半乳糖1-磷酸(取决于缺少哪种酶)的毒性水平降低,从而导致疾病状况,所述疾病状况包括但不限于:黄疸、肝(肿)大(增大的肝)、肾小管功能障碍、肌张力减退、败血症、肝细胞功能不全、肝硬化、肾衰竭、白内障、呕吐、癫痫、低血糖、嗜睡、脑损伤和卵巢衰竭。因此,当与对象中的半乳糖血症有关时,这些状况也可以通过本公开的方法适当地被治疗。

[0043] 根据对象中缺乏的酶,半乳糖血症可以被分为1型半乳糖血症(半乳糖-1-磷酸尿苷酰转移酶)、2型半乳糖血症(半乳糖激酶)和3型半乳糖血症(UDP半乳糖差向异构酶)。在一个实施方案中,本文公开的重组载体或方法可以被用于治疗1型、2型或3型半乳糖血症。在另一个实施方案中,重组体或方法可以被用于治疗1型半乳糖血症。因为半乳糖血症与多种疾病状况有关,在另一个实施方案中,本公开提供了一种在患有半乳糖血症的对象中减轻疾病状况的方法,其中所述疾病状况选自黄疸、肝(肿)大(增大的肝)、肾小管功能障碍、肌张力减退、败血症、肝细胞功能不全、肝硬化、肾衰竭、白内障、呕吐、癫痫、低血糖、嗜睡、脑损伤或卵巢衰竭。

[0044] 与半乳糖血症相关的其他疾病状况包括但不限于言语障碍、共济失调、弗里德希氏共济失调(Friedreich's Ataxia)、辨距不良、骨密度降低或卵巢功能早衰。这些状况也可以通过本公开的方法治疗。

[0045] 如本文使用的,术语“细胞”可以指原核或真核细胞,任选地从对象或可商购的来源获得。

[0046] “真核细胞”包含除无核原虫类之外的所有生命的界。通过膜结合的核可以很容易地区分它们。动物、植物、真菌和原生生物是真核生物或其细胞通过内膜和细胞骨架组织成复杂的结构的生物。最有特色的膜结合的结构是细胞核。除非特别说明,否则术语“宿主”包括真核宿主,包括例如酵母、高等植物、昆虫和哺乳动物细胞。真核细胞或宿主的非限制性

实施例包括猿猴、牛、猪、小鼠、大鼠、鸟、爬虫和人,例如,HEK293细胞和293T细胞。

[0047] “原核细胞”通常缺少细胞核或任何其他膜结合细胞器,被分为两类——细菌和古细菌。除了染色体DNA,这些细胞还可以在称为附加体的环状环中含有遗传信息。细菌细胞非常小,大约相当于动物线粒体的尺寸(直径约1-2  $\mu\text{m}$ ,长10  $\mu\text{m}$ )。原核细胞具有三种主要形状:杆状、球形和螺旋形。细菌细胞没有经历像真核生物这样复杂的复制过程,而是通过二分裂来分裂。实施例包括但不限于芽孢杆菌细菌、大肠杆菌细菌和沙门氏菌细菌。

[0048] 术语“编码”在被应用至核酸序列时是指,被陈述来“编码”多肽的多核苷酸,以其天然状态或者在通过本领域技术人员熟知的方法操纵时,可以被转录和/或翻译来产生用于该多肽和/或其片段的mRNA。反义链是这样的核酸的互补序列,并且编码序列可以由此导出。

[0049] 当涉及特定分子、生物或细胞材料时,术语“等效物”或“生物等效物”可互换使用,并且意指具有最小同源性且仍保持所需结构或功能的那些。等效多肽的非限制性实施例包括与其或与多肽序列具有至少60%、或者至少65%、或者至少70%、或者至少75%、或者80%、或者至少85%、或者至少90%、或者至少95%同一性的多肽,或由在高严格条件下与编码此类多肽序列的多核苷酸杂交的多核苷酸或其互补序列编码的多肽。高严格性条件在本文中被描述,并通过引用并入本文。替代地,其等效物是由与参考多核苷酸(例如野生型多核苷酸)具有至少70%、或者至少75%、或者80%、或者至少85%、或者至少90%、或者至少95%同一性、或者至少97%序列同一性的多核苷酸或其互补序列编码的多肽。

[0050] 等效多肽的非限制性实施例包括与参考多核苷酸具有至少60%、或者至少65%、或者至少70%、或者至少75%、或者80%、或者至少85%、或者至少90%、或者至少95%、或者至少97%同一性的多核苷酸。等效物还意指在高严格条件下与参考多核苷酸杂交的多核苷酸或其互补序列。

[0051] 多核苷酸或多核苷酸区域(或多肽或多肽区域)与另一个序列具有一定百分比(例如80%、85%、90%或95%)的“序列同一性”是指,当被比对时,该百分比的碱基(或氨基酸)在两个序列的比较中是相同的。可以使用本领域已知的软件程序来确定比对和同源性或序列同一性百分比,例如在Current Protocols in Molecular Biology (Ausubel et al., eds. 1987) Supplement 30, section 7.7.18, Table 7.7.1中描述的软件程序。在某些实施方案中,默认参数被用于比对。非限制性示例性比对程序是使用默认参数的BLAST。特别地,示例性程序包括BLASTN和BLASTP,使用以下默认参数:遗传密码 = 标准;筛选 = 无;链 = 两个;截点 = 60;预期 = 10;矩阵 = BLOSUM62;描述 = 50个序列;排序 = HIGH SCORE;数据库 = 不重复的,GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + SwissProtein + SPupdate + PIR。这些程序的详情可以在以下网址找到: [ncbi.nlm.nih.gov/cgi-bin/BLAST](http://ncbi.nlm.nih.gov/cgi-bin/BLAST)。序列同一性和同一性百分比可以通过将它们并入到clustalW(可从网址[genome.jp/tools/clustalw/](http://genome.jp/tools/clustalw/)获得,最后访问时间为2017年1月13日)中来确定。

[0052] “同源性”或“同一性”或“相似性”是指两个肽之间或两个核酸分子之间的序列相似性。可以通过比较为了进行比较而被比对的每个序列中的位置,确定同源性。当被比较的序列中的位置被相同的碱基或氨基酸占据时,则这些分子在该位置是同源的。序列之间的同源性的程度是这些序列享有的匹配的或同源的位置的数目的函数。“不相关的”或“非同

源的”序列与本文公开的序列中的一个享有少于40%的同一性、或者替代地少于25%的同一性。

[0053] “同源性”或“同一性”或“相似性”也可以指在严格条件下杂交的两个核酸分子。

[0054] “杂交”是指其中一个或多个多核苷酸反应来形成复合物,并且该复合物通过核苷酸残基的碱基之间的氢键结合而被稳定的反应。可以通过Watson-Crick碱基配对、Hoogsteen结合或通过任何其他序列特异性的方式来发生氢键结合。所述复合物可以包含形成双螺旋结构的两条链、形成多链复合物的三条或更多条链、单一的自我杂交的链、或这些的任何组合。杂交反应可以由在更广泛的过程中的步骤组成,例如PCR过程的起始步骤、或者通过核酶进行的多核苷酸的酶裂解步骤。

[0055] 严格杂交条件的实施例包括:约25°C至约37°C的孵育温度;约6x SSC至约10x SSC的杂交缓冲液浓度;约0%至约25%的甲酰胺浓度;以及约4x SSC至约8x SSC的洗涤溶液。中度杂交条件的实施例包括:约40°C至约50°C的孵育温度;约9x SSC至约2x SSC的缓冲液浓度;约30%至约50%的甲酰胺浓度;以及约5x SSC至约2x SSC的洗涤溶液。高严格杂交条件的实施例包括:约55°C至约68°C的孵育温度;约1x SSC至约0.1x SSC的缓冲液浓度;约55%至约75%的甲酰胺浓度;以及约1x SSC至约0.1x SSC的洗涤溶液、或去离子水。一般来说,杂交孵育时间为5分钟至24小时,具有1、2、或更多个洗涤步骤,并且洗涤孵育时间为约1、2或15分钟。SSC是0.15 M NaCl和15 mM柠檬酸缓冲液。应该理解的是,可以采用使用其他缓冲系统的SSC的等效物。

[0056] 如本文使用的,术语“表达”是指多核苷酸被转录成mRNA的过程和/或被转录的mRNA随后被翻译成肽、多肽或蛋白质的过程。如果多核苷酸衍生自基因组DNA,则表达可以包括mRNA在真核细胞中的剪接。

[0057] “基因”是指含有至少一个开放阅读框(ORF)的多核苷酸,其在转录和翻译后能够编码特定的多肽或蛋白质。“基因产物”或者“基因表达产物”是指当基因被转录和翻译时产生的氨基酸(例如,肽或多肽)。

[0058] “在转录控制之下”是本领域众所周知的术语,并且表示多核苷酸序列(通常是DNA序列)的转录取决于其被可操作地连接至有助于转录起始或促进转录的元件。“可操作地连接”意指以允许它们在细胞中起作用的方式排列多核苷酸。一方面,本发明提供了与下游序列可操作地连接的启动子,例如自杀基因、VEGF、165A VEGF、tet激活剂等。

[0059] 术语“编码”在被应用至多核苷酸时是指,被陈述来“编码”多肽的多核苷酸,以其天然状态或者在通过本领域技术人员熟知的方法操纵时,可以被转录和/或翻译来产生用于该多肽和/或其片段的mRNA。反义链是这样的核酸的互补序列,并且编码序列可以由此导出。

[0060] 如本文使用的,术语“分离的”是指,分子或生物体或细胞材料基本上不含其他材料。

[0061] 如本文使用的,术语“功能性”可以被用于修饰任何分子、生物或细胞材料,以使其实现特定的指定作用。

[0062] 如本文使用的,术语“核酸序列”和“多核苷酸”被可互换地使用来指任何长度的核苷酸(核糖核苷酸或脱氧核糖核苷酸)的聚合形式。因此,该术语包括但不限于单链、双链或多链DNA或RNA、基因组DNA、cDNA、DNA-RNA杂交体、或包含嘌呤和嘧啶碱基或其他天然的、化

学的或生物化学修饰的、非天然的或衍生的核苷酸碱基的聚合物。

[0063] 如本文使用的,术语“启动子”是指调控编码序列(例如基因)的表达的任何序列。例如,启动子可以是组成型的、诱导型的、抑制型的或组织特异性的。“启动子”是控制序列,其是多核苷酸序列中控制转录的起始和速率的区域。其可以包含调控蛋白和分子可以结合的遗传元件,例如RNA聚合酶和其他转录因子。非限制性示例性启动子包括Rous肉瘤病毒(RSV)、(任选地具有RSV增强子)、巨细胞病毒(CMV)启动子、SV40启动子、二氢叶酸还原酶启动子、 $\beta$ -肌动蛋白启动子、磷酸甘油激酶(PGK)启动子、U6启动子或EF1启动子。在一些实施方案中,启动子是鸡 $\beta$ -肌动蛋白(“CBA”)启动子(例如SEQ ID NO:2的编号为439至708的碱基对,或SEQ ID NO:3的编号为644至921的碱基对,或它们各自的等效物)。

[0064] 以下在本文中提供了具有某些靶标特异性的其他非限制性示例性启动子,包括但不限于CMV、EF1a、SV40、PGK1(人或小鼠)、P5、Ubc、人 $\beta$ 肌动蛋白、CAG、TRE、UAS、Ac5、多角体蛋白、CaMKIIa、Gal1、TEF1、GDS、ADH1、CaMV35S、Ubi、H1、U6和 $\alpha$ -1-抗胰蛋白酶启动子。合成来源的启动子可以被用于普遍存在的或组织特异性的表达。此外,上文提到的一些病毒来源的启动子(例如CMV、HIV、腺病毒和AAV启动子)可以被用于本文公开的方法。在一些实施方案中,启动子与增强子连接以增加转录效率。增强子的非限制性实施例包括RSV增强子或CMV增强子(例如,SEQ ID NO:2的编号为153至432的碱基对)。

[0065] 增强子是增加靶标序列表达的调控元件。“启动子/增强子”是含有能够提供启动子和增强子功能的序列的多核苷酸。例如,逆转录病毒的长末端重复序列含有启动子和增强子两者功能。增强子/启动子可以是“内源的”或“外源的”或“异源的”。“内源的”增强子/启动子是与基因组中给定基因天然连接的增强子/启动子。“外源的”或“异源的”增强子/启动子是通过基因操作(即分子生物学技术)的方式与基因并列放置的,使得该基因的转录由连接的增强子/启动子指导。

[0066] 术语“蛋白质”、“肽”和“多肽”被可互换地使用,并且在其最广泛的意义上是指两个或更多个氨基酸、氨基酸类似物或肽模拟物子单元的化合物。所述子单元可以通过肽键而被连接。在另一方面,所述子单元可以通过其他键(例如酯键、醚键等)被连接。蛋白质或肽必须包括至少两个氨基酸,并且对于可以组成蛋白质或肽的序列的氨基酸的最大数目没有限制。如本文使用的,术语“氨基酸”是指天然的和/或非天然的或合成的氨基酸,包括甘氨酸以及D和L光学异构体、氨基酸类似物和肽模拟物。

[0067] 如本文使用的,术语“载体”是指包含完整复制子的非染色体核酸,使得当被置于细胞内时,载体可以例如通过转化过程被复制。载体可以是病毒的或非病毒的。病毒载体包括逆转录病毒、腺病毒、疱疹病毒、杆状病毒、经修饰的杆状病毒、丘疹病毒或其他经修饰的天然病毒。用于递送核酸的示例性非病毒载体包括裸DNA、单独的或与阳离子聚合物组合的与阳离子脂质复合的DNA、阴离子和阳离子脂质体、在某些情况下包含在脂质体中的DNA-蛋白质复合物和包含与阳离子聚合物(如异质聚赖氨酸、定长寡肽和聚乙烯亚胺)缩合的DNA的颗粒、以及特定用途的包含病毒和聚赖氨酸-DNA的三元复合物。

[0068] “病毒载体”被定义为重组产生的病毒或病毒颗粒,其包含在体内、离体或体外递送至宿主细胞的多核苷酸。病毒载体的实施例包括逆转录病毒载体、AAV载体、慢病毒载体、腺病毒载体、甲病毒(Alphavirus)载体等。甲病毒载体(例如基于西门利克森林(Semliki Forest)病毒的载体和基于辛德毕斯(Sindbis)病毒的载体)也已被开发用于基因疗法和免

疫疗法。参见Schlesinger and Dubensky (1999) *Curr. Opin. Biotechnol.* 5:434-439 and Ying, et al. (1999) *Nat. Med.* 5(7):823-827。

[0069] 在另一个实施方案中,启动子是诱导型启动子。在一个特定的相关实施方案中,启动子是诱导型四环素启动子。Tet-Off和Tet-On基因表达系统使研究人员可以随时访问由Gossen&Bujard(1992; Tet-Off)和Gossen等人(1995; Tet-On)描述的受调控的高级基因表达系统。在Tet-Off系统中,当从培养基中去除四环素(Tc)或强力霉素(Dox, Tc衍生物)时,基因表达被打开。相反,通过添加Dox,可以在Tet-On系统中打开表达。两种系统都允许严格调控基因表达来应答Tc或Dox的浓度变化。Tet系统中的最高表达水平非常高,与可从强的组成型哺乳动物启动子(例如CMV)获得的最高表达水平相当(Yin等人,1996)。与其他诱导型哺乳动物表达系统不同,Tet系统中的基因调控是高特异性的,因此对结果的解释不会因多效性作用或非特异性诱导而变得复杂。在大肠杆菌中,Tet阻遏蛋白(TetR)负调控Tn10转座子上四环素抗性操纵子的基因。在不存在Tc的情况下,TetR通过与tet操纵子序列(tetO)结合来阻碍这些基因的转录。TetR和tetO为在哺乳动物实验系统中使用提供了调控和诱导的基础。在Tet-On系统中,调节蛋白基于由TetR中的四个氨基酸变化所产生的“反向”Tet阻遏子(rTetR)(Hillen & Berens, 1994; Gossen et al., 1995)。产生的蛋白质rtTA(反向tTA也称为四环素激活蛋白)由pTet-On调控质粒编码。该基因可以和GALT基因在分开的载体中,或在相同的基因上编码。

[0070] 在一个相关的实施方案中,载体还包含编码四环素激活蛋白的核酸以及调节四环素激活蛋白表达的启动子,或基本上由其组成,或进一步由其组成。

[0071] 在本文所述的载体、分离的细胞、病毒包装系统和方法中有用的其他诱导型系统包括通过蜕皮激素、雌激素、孕酮、二聚化的化学诱导剂和异丙基- $\beta$ -D1-硫代半乳糖苷(EPTG)的调节。

[0072] 如本文使用的,术语“重组表达系统”或“重组载体”是指用于通过重组形成的某些遗传物质的表达的一种或多种遗传构建体。

[0073] “基因递送载体”被定义为可以将插入的多核苷酸携带到宿主细胞中的任何分子。基因递送载体的实施例是脂质体、胶束生物相容的聚合物(包括天然聚合物和合成聚合物);脂蛋白;多肽;多糖;脂多糖;人工病毒包膜;金属颗粒;以及细菌、或病毒,例如杆状病毒、腺病毒和逆转录病毒、噬菌体、粘粒、质粒、真菌载体以及已经被描述用于在多种真核和原核宿主中表达的本领域通常使用的并且可以用于基因治疗以及简单的蛋白质表达的其他重组载体。

[0074] 可以使用基因递送载体将本文公开的多核苷酸递送至细胞或组织。如本文使用的,“基因递送”、“基因转移”、“转导”等是指将外源多核苷酸(有时称为“转基因”)引入宿主细胞的术语,而不论采用哪种方法引入。这样的方法包括多种众所周知的技术,例如载体介导的基因转移(通过,例如,病毒感染/转染,或各种其他基于蛋白质或基于脂质的基因递送复合物)以及促进“裸”多核苷酸递送的技术(例如电穿孔,“基因枪(gene gun)”递送和用于引入多核苷酸的各种其他技术)。所引入的多核苷酸可以在宿主细胞中稳定或瞬时维持。稳定的维持通常需要所引入的多核苷酸包含与宿主细胞相容的复制起点或整合到宿主细胞的复制子中,例如染色体外复制子(例如质粒)或核或线粒体染色体。如本领域已知和本文所述的,已知许多载体能够介导基因向哺乳动物细胞的转移。

[0075] “质粒”是与染色体DNA分离的染色体外DNA分子,其能够独立于染色体DNA而进行复制。在许多情况下,它是环形的和双链的。质粒提供了在微生物群体内水平基因转移的机制,并且通常在给定的环境状态下提供选择优势。质粒可以携带在竞争性环境中对天然存在的抗生素具有抗性的基因,或者产生的蛋白质可以在类似情况下充当毒素。

[0076] 基因工程中使用的“质粒”被称为“质粒载体”。许多从商业上获得的质粒可用于此类用途。要复制的基因被插入含有使细胞对特定抗生素具有抗性的基因和多个克隆位点(MCS,或多连接子)的质粒的拷贝中,所述克隆位点是含有几个常用的限制性酶切位点的短区域,其允许在该位置轻松插入DNA片段。质粒的另一主要用途是制备大量蛋白质。在这种情况下,研究人员使含有带有目的基因的质粒的细菌生长。正如细菌产生蛋白质以赋予其抗生素抗性一样,其也可以被诱导以从插入的基因产生大量蛋白质。

[0077] 在基因转移由DNA病毒载体(例如腺病毒(Ad)或腺相关病毒(AAV))介导的方面中,载体构建体是指包含病毒基因组或其一部分和转基因的多核苷酸。腺病毒(Ad)是相对良好地被表征的同质病毒群,包括超过50种血清型。Ad不需要整合到宿主细胞基因组中。还构建了重组Ad衍生的载体,特别是那些减少野生型病毒的重组和产生的可能性的载体。这类载体可以从例如Takara Bio USA (Mountain View, CA)、Vector Biolabs (Philadelphia, PA)和Creative Biogene (Shirley, NY)的来源商购获得。野生型AAV具有整合到宿主细胞的基因组中的高感染性和特异性。参见,Wold and Toth (2013) *Curr. Gene. Ther.* 13 (6):421-433, Hermonat & Muzyczka (1984) *Proc. Natl. Acad. Sci. USA* 81:6466-6470和Lebkowski et al. (1988) *Mol. Cell. Biol.* 8:3988-3996。

[0078] 含有启动子和可以被可操作地连接进多核苷酸的克隆位点的载体是本领域众所周知的。这样的载体能够在体外或体内转录RNA,并且可以商购获得,例如从Agilent Technologies (Santa Clara, Calif.)和Promega Biotech (Madison, Wis.)。为了优化表达和/或体外转录,在转录或翻译水平上,可能有必要去除、添加或改变克隆的5'和/或3'非翻译部分,以消除可能干扰或减少表达的额外的、潜在的不适当的替代翻译起始密码子或其他序列。或者,可以在起始密码子的5'处立即插入共有核糖体结合位点以增强表达。

[0079] 基因递送载体还包括DNA/脂质体复合物、胶束和靶向病毒蛋白-DNA复合物。还包含靶向抗体或其片段的脂质体可以被用于本文公开的方法中。除了将多核苷酸递送到细胞或细胞群外,通过非限制性的蛋白质转染技术可以将本文公开的蛋白质直接引入细胞或细胞群,替代地能够增强本文公开的蛋白质的表达和/或促进活性的培养条件是其他非限制性技术。

[0080] 如本文使用的,术语“信号肽”或“信号多肽”意指通常存在于新合成的分泌或膜多肽或蛋白质的N-末端的氨基酸序列。它起到将多肽指导至特定的细胞位置的作用,例如穿过细胞膜、进入细胞膜或进入细胞核。在一些实施方案中,在定位之后信号肽被移除。信号肽的实施例是本领域熟知的。非限制性实施例是美国专利号8,853,381、5,958,736和8,795,965中描述的那些。

[0081] 如本文使用的,术语“病毒衣壳”或“衣壳”是指病毒颗粒的蛋白质壳或外壳。衣壳的功能是衣壳化(encapsidate)、保护、运输病毒基因组并将其释放到宿主细胞中。衣壳通常由蛋白质(“衣壳蛋白”)的寡聚结构亚基组成。如本文所用,术语“衣壳化”是指封闭在病毒衣壳内。

[0082] 如本文使用的,关于病毒或质粒的术语“辅助(子)”是指用于提供复制和包装病毒颗粒或重组病毒颗粒(例如本文公开的经修饰的AAV)所必需的另外的组分的病毒或质粒。辅助病毒编码的组分可以包括病毒粒子组装、衣壳化、基因组复制和/或包装所需的任何基因。例如,辅助病毒可以编码用于病毒基因组的复制的必需的酶。辅助病毒和适用于AAV构建体的质粒的非限制性实施例包括pHELP(质粒)、腺病毒(病毒)或疱疹病毒(病毒)。

[0083] 如本文使用的,术语“AAV”是腺相关病毒的标准缩写。腺相关病毒是一种单链DNA细小病毒,仅在通过共感染辅助病毒提供某些功能的细胞中生长。AAV的一般信息和评论可以参见例如Carter, 1989, Handbook of Parvoviruses, Vol. 1, pp. 169- 228, and Berns, 1990, Virology, pp. 1743-1764, Raven Press, (New York)。完全可以预期,这些评论中描述的相同原理将适用于在评论公布之日后表征的其他AAV血清型,因为众所周知,各种血清型在结构上和功能上甚至在遗传水平上都十分紧密相关。(参见,例如, Blacklowe, 1988, pp. 165-174 of Parvoviruses and Human Disease, J. R. Pattison, ed.; and Rose, Comprehensive Virology 3: 1-61 (1974))。例如,所有AAV血清型显然都表现出由同源rep基因介导的非常相似的复制特性;并且都携带有三种相关的衣壳蛋白,例如在AAV2中表达的那些。异源双链分析进一步表明了相关程度,该分析揭示了沿基因组长度的血清型之间广泛的交叉杂交;以及在末端与“反向末端重复序列”(ITR)相对应的类似自退火片段的的存在。相似的传染性模式也表明,每种血清型中的复制功能都处于相似的调节控制之下。

[0084] 如本文使用的,“AAV载体”是指包含侧翼为AAV末端重复序列(ITR)的一个或多个感兴趣的多核苷酸(或转基因)的载体。当存在于已经用编码和表达rep和cap基因产物的载体转染的宿主细胞中时,这样的AAV载体可以被复制并被包装成感染性病毒颗粒。

[0085] “AAV病毒粒子”或“AAV病毒颗粒”或“AAV载体颗粒”是指由至少一种AAV衣壳蛋白和衣壳化的多核苷酸AAV载体组成的病毒颗粒。如果颗粒包含异源多核苷酸(即,野生型AAV基因组以外的多核苷酸,例如要递送至哺乳动物细胞的转基因),则通常称为“AAV载体颗粒”或简称为“AAV载体”。因此,AAV载体颗粒的产生必然包括AAV载体的产生,因为这种载体被包含在AAV载体颗粒内。

[0086] 在一些实施方案中,AAV是一种复制缺陷型细小病毒,其单链DNA基因组的长度约为4.7 kb,包括两个145个核苷酸的反向末端重复序列(ITR)。AAV有多种血清型。AAV血清型的基因组的核苷酸序列是已知的。例如,GenBank登录号NC\_002077中提供了AAV-1的完整基因组;GenBank登录号NC\_001401和Srivastava et al., J. Virol., 45: 555-564 (1983)中提供了AAV-2的完整基因组;GenBank登录号NC\_1829中提供了AAV-3的完整基因组;GenBank登录号NC\_001829中提供了AAV-4的完整基因组;GenBank登录号AF085716中提供了AAV-5的完整基因组;GenBank登录号NC\_001862中提供了AAV-6的完整基因组;GenBank登录号AX753246和AX753249中分别提供了AAV-7和AAV-8基因组的至少一部分;Gao et al., J. Virol., 78: 6381-6388 (2004)中提供了AAV-9基因组;Mol. Ther., 13(1): 67-76 (2006)中提供了AA-10基因组;以及Virology, 330(2): 375-383 (2004)中提供了AAV-11基因组。美国专利9,434,928中提供了AAV rh.74基因组的序列,其内容通过引用并入本文。美国专利号9,434,928还提供衣壳蛋白和自互补基因组的序列。在一方面,基因组是自互补基因组。指导病毒DNA复制(rep)、衣壳化/包装和宿主细胞染色体整合的顺式作用序列被包

含在AAV ITR中。三个AAV启动子(因其相对图谱位置而命名为p5、p19和p40)驱动编码rep和cap基因的两个AAV内部开放阅读框的表达。两个rep启动子(p5和pi9),加上单个AAV内含子的差异剪接(在核苷酸2107和2227处),导致从rep基因产生四个rep蛋白(rep78、rep68、rep52和rep40)。Rep蛋白具有最终负责复制病毒基因组的多种酶学性质。Cap基因是从p40启动子表达的,并且它编码三个衣壳蛋白VP1、VP2和VP3。替代的剪接和非共有的翻译起始位点负责三个相关的衣壳蛋白的产生。单个共有多聚腺苷酸化位点位于AAV基因组的图谱位置95。Muzyczka, *Current Topics in Microbiology and Immunology*, 158: 97-129 (1992)中综述了AAV的生命周期和遗传学。

[0087] AAV具有独特的功能,使其作为将外源DNA递送至细胞的载体(例如在基因治疗中)有吸引力。培养物中细胞的AAV感染是非细胞病变的,人类和其他动物的自然感染是隐形的且无症状的。此外,AAV感染许多哺乳动物细胞,从而有可能在体内靶向许多不同的组织。此外,AAV缓慢转导分裂和非分裂的细胞,并且可以作为转录活性核附加体(染色体外元素)在这些细胞的生命周期中持续存在。AAV前病毒基因组作为克隆的DNA被插入质粒中,这使得重组基因组的构建成为可能。此外,由于指导AAV复制和基因组衣壳化的信号被包含在AAV基因组的ITR中,所以基因组内部大约4.3 kb的部分或全部(编码复制和结构衣壳蛋白,rep-cap)可以被外源DNA代替。为了产生AAV载体,可以以反式形式提供rep和cap蛋白。AAV的另一个重要特征是它是一种非常稳定和强烈(hearty)的病毒。它可以轻松承受用于灭活腺病毒的条件(56℃至65℃持续数小时),从而使AAV的冷藏不那么关键。AAV甚至可以冻干。最后,AAV感染的细胞对超级感染没有抵抗力。

[0088] 多项研究证明,重组AAV介导的蛋白质在肌肉中长期表达(>1.5年)。参见,Clark et al., *Hum Gene Ther*, 8: 659-669 (1997); Kessler et al., *Proc Natl Acad Sci USA*, 93: 14082-14087 (1996)和Xiao et al., *J Virol*, 70: 8098-8108 (1996)。还参见,Chao et al., *Mol Ther*, 2:619-623 (2000)和Chao et al., *Mol Ther*, 4:217-222 (2001)。此外,如在Herzog et al., *Proc Natl Acad Sci USA*, 94: 5804-5809 (1997)和Murphy et al., *Proc Natl Acad Sci USA*, 94: 13921- 13926 (1997)中所述,由于肌肉高度血管化,所以重组AAV转导导致肌肉注射后全身循环中出现转基因产物。此外,Lewis et al., *J Virol*, 76: 8769-8775 (2002)证明了骨骼肌纤维具有正确的抗体糖基化、折叠和分泌所必需的细胞因子,表明肌肉能够稳定表达分泌的蛋白质治疗剂。本发明的重组AAV(rAAV)基因组包含编码GALT的核酸分子(例如,SEQ ID NO:1)和在该核酸分子侧翼的一个或多个AAV ITR。rAAV基因组中的AAV DNA可以来自用于可以衍生包括但不限于以下AAV血清型的重组病毒的任何AAV血清型:AAV-1、AAV-2、AAV-3、AAV-4、AAV-5、AAV-6、AAV-7、AAV-8、AAV-9、AAV-10、AAV-11、AAV-12、AAV-13、AAV PHP.B和AAV rh74。假型rAAV的产生公开于例如WO 01/83692。也可以预期其他类型的rAAV变体,例如具有衣壳突变的rAAV。参见,例如Marsic et al., *Molecular Therapy*, 22(11): 1900-1909 (2014)。各种AAV血清型的基因组的核苷酸序列是本领域已知的。

[0089] 如本文使用的,关于病毒衣壳蛋白的术语“外部”是指在组装病毒衣壳中面向外部的衣壳蛋白的表面、结构域、区域或末端。关于病毒衣壳蛋白的术语“内部”是指在组装病毒衣壳中面向内部的衣壳蛋白的表面、结构域、区域或末端(氨基末端或羧基末端)。当参考组装病毒衣壳使用时,术语“内部”是指病毒衣壳内部的衣壳化空间以及暴露于封闭空间的衣

壳的面向内的表面。内部空间被病毒衣壳蛋白衣壳化,并且可以包含核酸(例如病毒基因组、病毒蛋白、宿主或包装细胞的蛋白)以及在复制、病毒粒子组装、衣壳化和/或包装过程中包装或衣壳化的任何其他组分或因子。

[0090] 如本文使用的,术语“缀合”是指将病毒衣壳蛋白附接、偶联、融合和/或连接至GALT蛋白或其等效物的任何方法。缀合的非限制性实施例包括:重组融合蛋白,其中GALT蛋白或其等效物和病毒衣壳蛋白由包含GALT蛋白或其等效物和病毒衣壳蛋白的基因的单个多核苷酸编码;基于模块化内含肽的GALT-内含肽蛋白和病毒衣壳-内含肽蛋白的组装;翻译后修饰,其引起GALT蛋白或其等效物与病毒衣壳蛋白之间形成化学键;以及GALT或其等效物与病毒衣壳蛋白通过一个或多个连接子连接。在一些实施方案中,缀合可以是病毒衣壳蛋白与其等效物之间的暂时或短暂结合状态。例如,GALT或其等效物可以通过在感染后期或特定细胞微环境内对pH或离子梯度变化敏感的聚合物与病毒衣壳蛋白瞬时连接,例如肟键(参见,例如Jin et al. *Biomacromolecules*, 2011, 12 (10), pp 3460-3468 and Yoshida et al. *Expert Opin Drug Deliv.* 2013 Nov; 10(11): 1497-1513)。

[0091] 如本文所用,术语“标记”意指直接或间接可检测的化合物或组合物直接或间接缀合至待检测的组合物,例如多核苷酸或蛋白质如抗体以生成“标记的”成分。该术语还包括与多核苷酸缀合的序列,该序列将在插入序列表达时提供信号,例如绿色荧光蛋白(GFP)等。标记本身可以是可检测的(例如,放射性同位素标记或荧光标记),或者在酶标记的情况下,可以催化可检测的底物化合物或组合物的化学变化。标签可适合于小规模检测,或更适合于高通量筛选。这样,合适的标记包括但不限于放射性同位素、荧光染料、化学发光化合物、染料和蛋白质(包括酶)。标记可以被简单地检测或可以被量化。被简单检测到的响应通常包括仅被确认存在的响应,而被量化的响应通常包括具有诸如强度、极化和/或其他性质的可量化(例如可数字报告)值的响应。在发光或荧光测定中,可直接使用与实际参与结合的测定组分相关的发光体或荧光团直接产生可检测的响应,或间接使用与另一(例如报道分子或指示剂)组分相关的发光体或荧光团间接产生可检测的响应。

[0092] 产生信号的发光标记的实施例包括但不限于生物发光和化学发光。可检测的发光响应通常包括发光信号的改变或出现。用于发光标记测定组分的合适方法和发光体是本领域已知的,例如在Haugland, Richard P. (1996) *Handbook of Fluorescent Probes and Research Chemicals* (第六版)中进行了描述。发光探针的实施例包括但不限于水母发光蛋白和萤光素酶。

[0093] 合适的荧光标记的实施例包括但不限于荧光素、罗丹明、四甲基罗丹明、曙红、赤藓红、香豆素、甲基香豆素、芘、孔雀石绿、二苯乙烯、荧光素黄、级联蓝™和得克萨斯红。其他合适的光学染料在Haugland, Richard P. (1996) *Handbook of Fluorescent Probes and Research Chemicals* (第6版)中进行了描述。

[0094] 在另一方面,对荧光标记进行功能化以促进共价附接到细胞或组织的表面中或表面上存在的细胞成分(例如细胞表面标记)。合适的官能团包括但不限于异硫氰酸酯基、氨基、卤代乙酰基、马来酰亚胺、琥珀酰亚胺酯和磺酰基,所有这些均可以用于将荧光标记附着于第二分子。荧光标记的官能团的选择将取决于与连接子、试剂、标记或第二标记试剂的连接位点。

[0095] 荧光标记的连接可以直接与细胞组分或化合物连接,或者可以通过连接子连接。

用于将荧光标记间接连接至中间体的合适的结合对包括但不限于抗原/抗体,例如罗丹明/抗罗丹明、生物素/抗生物素蛋白和生物素/链霉亲和素。

[0096] 短语“固体支持物”是指非水相表面,例如“培养板”、“基因芯片”或“微阵列”。这样的基因芯片或微阵列可以通过本领域技术人员已知的多种技术用于诊断和治疗目的。在一种技术中,寡核苷酸被连接并排列在基因芯片上,以通过杂交方法(例如美国专利号6,025,136和6,018,041中概述的方法)确定DNA序列。本发明的多核苷酸可以被修饰成探针,该探针又可以被用于检测遗传序列。例如在美国专利号5,968,740和5,858,659中描述了这样的技术。探针也可以连接或固定至电极表面,用于核酸序列的电化学检测,例如Kayem等人美国专利号5,952,172和Kelley et al. (1999) *Nucleic Acids Res.* 27:4830-4837中描述的。

[0097] “组合物”旨在表示活性多肽、多核苷酸或抗体与惰性(例如,可检测的标记)或活性(例如,基因递送载体)的另一种化合物或组合物的组合。

[0098] “药物组合物”旨在包括活性多肽、多核苷酸或抗体与惰性或活性载体(例如固体支持物)的组合,从而使该组合物适合于体外、体内或离体的诊断或治疗用途。

[0099] 如本文使用的,术语“药学上可接受的载体”包含任何标准药物载体,例如磷酸盐缓冲盐溶液、水和乳剂(例如油/水或水/油乳剂),以及各种类型的润湿剂。该组合物还可以包含稳定剂和防腐剂。载体、稳定剂和佐剂的实施例参见Martin (1975) *Remington's Pharm. Sci.*, 15th Ed. (Mack Publ. Co., Easton)。

[0100] 诊断或治疗的“对象”是细胞或动物,例如哺乳动物或人类。对象不仅限于特定物种,还包括经受诊断或治疗的非人类动物,并且是经受感染或动物建模的,例如猿猴、鼠类(例如大鼠、小鼠、灰鼠)、犬类(例如狗)、兔类(例如兔子)、家畜、运动动物和宠物。该术语也包括人类患者。

[0101] 术语“组织”在本文中用于指活的或死的生物的组织,或衍生自或设计为模仿活的或死的生物的任何组织。该组织可以是健康的、患病的和/或具有遗传突变。生物组织可以包括任何单个组织(例如,可以相互连接的细胞的集合)或组成生物体的器官或部分或区域的一组组织。组织可以包含均质的细胞材料,或者可以是复合结构,例如在包括胸部的身体区域中发现的复合结构,所述胸部例如可以包括肺组织、骨骼组织和/或肌肉组织。示例性组织包括但不限于源自肝脏、肺、甲状腺、皮肤、胰腺、血管、膀胱、肾脏、脑、胆管树、十二指肠、腹主动脉、静脉、心脏和肠的组织,包括其任何组合。

[0102] 如本文使用的,“治疗”对象中的疾病是指(1)防止症状或疾病在预先有倾向的或尚未显示疾病症状的对象中发生;(2)抑制疾病或阻止其发展或复发;或者(3)改善或消退疾病或疾病的症状。如本领域中理解的,“治疗”是用于获得有益的或期望的结果(包括临床结果)的方法。对于本技术的目的,有益的或期望的结果可以包括但不限于以下中的一种或多种:一个或多个症状的缓解或改善,病症(包括疾病)的程度的降低,病症(包括疾病)的稳定化(即不恶化)状态,病症(包括疾病)的延迟或减缓,病症(包括疾病)、状态和缓解(无论是部分还是全部)的进展、改善或缓解,无论是可检测的还是不可检测的。

[0103] 如本文使用的,术语“有效量”旨在表示足以实现期望效果的量。在治疗或预防应用的情况下,有效量将取决于所讨论病症的类型和严重性以及个体对象的特征,例如总体健康、年龄、性别、体重和对药物组合物的耐受性。在基因疗法的情况下,在一些实施方案

中,有效量是足以导致对象中缺乏的基因的部分或全部功能恢复的量。在其他实施方案中,AAV病毒颗粒的有效量是足以导致对象中基因的表达的量。在一些实施方案中,有效量是有需要的对象中增加半乳糖代谢所需的量。技术人员将能够根据这些和其他因素确定适当的量。

[0104] 在一些实施方案中,有效量将取决于所讨论的应用的大小和性质。它还将取决于靶标对象的性质和敏感性以及所使用的方法。本领域技术人员将能够基于这些和其他考虑因素确定有效量。根据实施方案,有效量可以包括组合物的一次或多次施用。

[0105] 如本文使用的,术语“给药”或“施用”旨在表示将物质递送至对象例如动物或人。施用可以在整个治疗过程中以一个剂量连续或间歇地进行。确定最有效的施用方式和剂量的方法是本领域技术人员已知的,并且将根据用于治疗的组合物、治疗的目的、以及所治疗的对象的年龄、健康状况或性别而变化。可以通过治疗医师或者对于宠物或动物通过治疗兽医,选择剂量水平和方式进行单次或多次施用。合适的剂型和施用试剂的方法是本领域已知的。也可以确定施用途径,并且确定最有效施用途径的方法是本领域技术人员已知的,并且将根据用于治疗的组合物、治疗目的、被治疗的对象的健康状况或疾病阶段、以及靶细胞或组织而变化。施用途径的非限制性实例包括静脉内、动脉内、肌肉内、心脏内、囊内、脑室内、硬膜外、大脑内、脑室内、视网膜下、玻璃体内、关节内、眼内、腹膜内、子宫内、皮肤内、皮下、经皮、经粘膜和吸入。

[0106] 执行本公开技术的方式

AAV载体、衣壳和制备方法

本文提供了一种重组多核苷酸或腺相关病毒(“AAV”)载体,其包含编码半乳糖-1-磷酸尿苷酰转移酶(“GALT”)的多核苷酸序列,或基本上由其组成,或进一步由其组成。在一方面,编码GALT的多核苷酸序列包含与SEQ ID NO:1至少85%、或90%、或95%、或97%、或99%相同的核苷酸序列,或基本上由其组成,或进一步由其组成。在一方面,序列与SEQ ID NO:1至少85%、或90%、或95%、或97%、或99%相同,条件是已经从野生型序列中被修饰的核苷酸中的至少一个、或两个、或三个、或四个、或五个、或六个、或七个、或八个、或九个、或十个或更多个、或全部没有从SEQ ID NO:1中被修饰。在另一方面,编码GALT的多核苷酸序列包含SEQ ID NO:1所示的核苷酸序列,或基本上由其组成,或进一步由其组成。在另一个实施方案中,多核苷酸序列编码SEQ ID NO:4的氨基酸序列或其等效物。

[0107] 在一方面,载体是自互补载体或单链DNA(ssDNA)载体。

[0108] 在一方面,AAV包含AAV(例如侧翼为两个反向末端重复序列(ITR)的scAAV或ssAAV),或基本上由其组成,或进一步由其组成。这些ITR在序列的末端形成发夹来作为引物,从而在随后的感染步骤开始之前开始第二链的合成。第二链合成被认为是有效感染的几个阻碍之一。scAAV的其他优点包括体外和体内转基因表达的增加和延长。因此,在一方面,AAV还包含两个ITR。

[0109] 用于产生载体的重组AAV骨架的非限制性实施例包括选自以下血清型的AAV载体:AAV1、AAV2、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、AAV13、AAV PHP.B或AAV rh74。在另一方面,载体骨架是AAV9血清型、rh74血清型或经修饰的AAVrh74血清型。

[0110] 编码GALT的多核苷酸序列任选地可操作地连接至启动子、组织特异性控制元件或组成型启动子。组成型启动子的非限制性实施例包括,例如,Rous肉瘤病毒(RSV)LTR启动子

(任选地具有RSV增强子)、巨细胞病毒(CMV)启动子、SV40启动子、二氢叶酸还原酶启动子、 $\beta$ -肌动蛋白启动子、磷酸甘油激酶(PGK)启动子或EF1启动子。在一个特定方面, $\beta$ -肌动蛋白启动子是鸡 $\beta$ -肌动蛋白(“CBA”)启动子。在另一个实施方案中,启动子选自CMV启动子、EF1a启动子、SV40启动子、PGK1(人或小鼠)启动子、P5启动子、Ubc启动子、人 $\beta$ 肌动蛋白启动子、CAG启动子、TRE启动子、UAS启动子、Ac5启动子、多角体蛋白启动子、CaMKIIa启动子、Gal1启动子、TEF1、GDS启动子、ADH1启动子、CaMV35S启动子、Ubi启动子、H1启动子、U6启动子或 $\alpha$ -1-抗胰蛋白酶启动子。

[0111] 在另一方面,重组多核苷酸或AAV进一步包含编码增强子元件的多核苷酸。非限制性实施例包括CMV增强子、WPRE和RSV增强子。

[0112] 在一个特定的方面,重组多核苷酸或AAV载体包含与SEQ ID NO:2或SEQ ID NO:3至少85%、或90%、或95%、或97%或99%相同的核苷酸序列,或基本上由其组成,或进一步由其组成。在一方面,序列与SEQ ID NO:2或3至少85%、或90%、或95%、或97%或99%相同,条件是已经从野生型序列中被修饰的核苷酸中的至少一个、或两个、或三个、或四个、或五个、或六个、或七个、或八个、或九个、或十个或更多个、或全部没有从SEQ ID NO:2或3中被修饰。在另一方面,重组AAV载体包含SEQ ID NO:2或SEQ ID NO:3所示的核苷酸序列,或基本上由其组成,或进一步由其组成。

[0113] 如本文所述的重组多核苷酸或AAV载体可以进一步包含可检测的或纯化标记。多核苷酸和/或载体可以作为包含媒介和载体(例如防腐剂或药学上可接受的载体)的组合物而被包含。

[0114] 目前,AAV载体的递送依赖于基于病毒的自然趋向性的将血清型选择用于组织靶向或直接注射入靶标组织。如果需要全身递送来获得最大的治疗益处,那么血清型选择是结合组织特异性启动子用于组织靶向的唯一可用选择。因此,AAV载体可以被包装到具有组织趋向性的衣壳中。

[0115] 在一些方面,载体被封装在野生型或经修饰的衣壳颗粒的病毒衣壳中。在一方面,本公开提供衣壳蛋白、分离的多核苷酸、制备衣壳蛋白的方法、重组病毒颗粒和用于产生病毒颗粒的重组表达系统。在一个实施方案中,病毒衣壳蛋白包含野生型衣壳蛋白或者通过氨基酸置换或插入1至7个氨基酸来修饰的病毒衣壳蛋白,或基本上由其组成,或进一步由其组成。在一些实施方案中,病毒衣壳蛋白是VP1,任选地是AAV9、AAV PHP.B或经修饰的AAVrh74。例如,AAV PHP.B具有AAV9 VP1的498位经修饰的氨基酸(从天冬酰胺到赖氨酸),以减少肝脏趋向性。在进一步的实施方案中,修饰包含在AAVrh74的VP1的502位氨基酸处异亮氨酸取代天冬酰胺或等效修饰。在一些实施方案中,修饰包含在AAVrh74的VP1的氨基酸505处色氨酸被取代为精氨酸。在一些实施方案中,修饰是在AAVrh74的VP1的591位氨基酸处插入肽YIG或YIGSR。在一些实施方案中,所述肽对 $\alpha$ 7 $\beta$ 1整合蛋白具有高亲和力和/或位于可能改变正常rh74受体结合的区域中。本发明还提供了这些多肽的等效物和编码它们的多核苷酸,其中所述等效物具有至少85%、或90%、或95%、或97%、或99%序列同一性,条件是氨基酸和多核苷酸中的一个或多个、两个或更多个、三个或更多个、四个或更多个、五个或更多个、六个或更多个、七个或更多个、或全部没有从突变序列中发生改变。

[0116] 可以使用病毒包装系统(例如逆转录病毒、腺病毒、疱疹病毒或杆状病毒包装系统)包装病毒(例如AAV)。在一些实施方案中,通过使用辅助病毒或辅助质粒和细胞系来实

现包装。辅助病毒或辅助质粒含有促进遗传物质向细胞内传递的成分和序列。在另一方面，将辅助质粒或包含辅助质粒的多核苷酸稳定地并入包装细胞系的基因组中，使得包装细胞系不需要用辅助质粒进行另外的转染。

[0117] 辅助质粒可以包含例如至少一个病毒辅助DNA序列，其衍生自反式编码包装复制缺陷型AAV所需的所有病毒粒子蛋白的复制缺陷型病毒基因组，用于产生能够以高滴度包装复制缺陷型AAV的病毒粒子蛋白，而不产生有复制能力的AAV。病毒DNA序列缺少编码病毒的病毒性5' LTR的天然增强子和/或启动子的区域，并且缺少负责包装辅助基因组的psi功能序列和3' LTR，但是编码外来的聚腺苷酸化位点（例如SV40聚腺苷酸化位点）、以及引导其中需要病毒生产的细胞类型中的有效转录的外来增强子和/或启动子。病毒是白血病毒，例如莫洛尼鼠白血病毒（MMLV）、人类免疫缺陷病毒（HIV）、或长臂猿白血病毒（GALV）。

[0118] 外来增强子和/或启动子可以是人类巨细胞病毒（HCMV）即时早期（IE）增强子和启动子、莫洛尼鼠肉瘤病毒（MMSV）的增强子和启动子（U3区域）、Rous肉瘤病毒（RSV）的U3区域、脾病灶形成病毒（SFFV）的U3区域、或连接到天然莫洛尼鼠白血病毒（MMLV）启动子的HCMV IE增强子。辅助质粒可以由两个逆转录病毒辅助DNA序列组成，它们由基于质粒的表达载体编码，例如其中第一辅助序列包括编码嗜亲性MMLV或GALV的gag和pol蛋白的cDNA，并且第二辅助序列包括编码env蛋白的cDNA。Env基因确定宿主范围，可以衍生自编码以下的基因：嗜异性的、双嗜性的、嗜亲性、多嗜性的（貂病灶形成）或10A1鼠白血病毒env蛋白、或长臂猿白血病毒（GALV）env蛋白、人类免疫缺陷病毒env（gp160）蛋白、水泡性口炎病毒（VSV）G蛋白、人类T细胞白血病（HTLV）I型和II型env基因产物，嵌合包膜基因，其衍生自前述env基因或编码前述env基因产物的细胞质和跨膜结构域的嵌合包膜基因中的一个或多个以及针对在所需的靶细胞上的特异性表面分子的单克隆抗体的组合。

[0119] 在包装过程中，将辅助质粒和编码AAV病毒蛋白的质粒瞬时共转染到能够产生病毒的哺乳动物细胞（例如人胚胎肾细胞，例如293细胞（ATCC No. CRL1573，ATCC，Rockville，Md.））的第一群体中，以产生高滴度的含重组逆转录病毒的上清液。在本发明的另一种方法中，然后将该瞬时转染的第一细胞群与哺乳动物靶细胞共培养，以用外源基因高效转导靶细胞。

#### [0120] 包装系统

本发明还提供一种病毒包装系统，其包含：如上所述的载体，其中骨架衍生自质粒、病毒；包装质粒；和包膜质粒。包装质粒含有核苷、衣壳和基质蛋白。专利文献中也描述了包装质粒的实施例，例如美国专利号7,262,049、6,995,258、7,252,991和5,710,037，其内容通过引用并入本文。该系统还含有编码由包膜质粒提供的假型包膜蛋白的质粒。假型病毒载体由携带衍生自其他包膜病毒的糖蛋白或者包含功能部分的载体颗粒组成。参见，例如美国专利号7,262,049，其内容通过引用并入本文。在一个优选的方面，包膜质粒编码不引起病毒颗粒非特异性结合细胞或细胞群的包膜蛋白。病毒颗粒的特异性由插入颗粒中的抗体结合结构域赋予。合适的包膜蛋白的实施例包括但不限于含有金黄色葡萄球菌ZZ结构域的那些。用于包膜的糖蛋白的选择部分取决于与颗粒缀合的抗体。

[0121] 本公开还提供了合适的包装细胞系。一方面，包装细胞系是HEK-293细胞系。其他合适的细胞系是本领域已知的，例如，在美国专利号7,070,994、6,995,919、6,475,786、6,372,502、6,365,150和5,591,624的专利文献中描述的，每一篇的内容通过引用并入本文。

[0122] 本发明进一步提供了一种在适合包装病毒载体的条件下产生假型AAV颗粒的方法,该方法包含用如上所述的病毒系统转导包装细胞系,或基本上由其组成,或进一步由其组成。这样的条件是本领领域已知的,并且在本文中进行了简要描述。可以使用本领领域技术人员已知的方法,例如离心,从细胞上清液中分离假型病毒颗粒。本发明进一步提供了这种分离的颗粒。

[0123] 本发明进一步提供了通过该方法产生的分离的多核苷酸和/或AAV病毒颗粒。假型病毒颗粒包含如本文所述的且编码GALT蛋白或其等效物(例如,如上文所述的SEQ ID NO. 4或SEQ ID NO. 4的等效物)的多核苷酸,或基本上由其组成,或进一步由其组成。

[0124] 分离的假型颗粒可以与保留结合预选细胞受体能力的抗体或抗体片段(例如,包含至少Fc结构域的片段)中的一种或多种缀合。

[0125] 抗体不是物种特异性的。换句话说,抗体可以是多克隆的或单克隆的,并且可以是小鼠、羊、人或其他物种。另外,它们可以是嵌合的或人源化的。

[0126] 宿主细胞

本发明还进一步提供了分离的细胞或细胞群,其包含如上所述的和通过引用并入本文的分离的多核苷酸、病毒颗粒、载体和包装系统,或基本上由其组成,或进一步由其组成。一方面,分离的细胞是包装细胞系。

[0127] 本发明还提供了分离的细胞或细胞群,其包含编码如本文所述的GALT蛋白或其等效物的多核苷酸序列和调节编码GALT的核酸表达的组成型或诱导型启动子,或基本上由其组成,或进一步由其组成。在一个实施方案中,启动子是如本文所述的诱导型启动子。在另一方面,启动子是如本文所述的组成型启动子。在另一方面,编码GALT的核酸包含SEQ ID NO.:1或其等效物,或基本上由其组成,或进一步由其组成。在一方面,该序列与SEQ ID NO.:1至少85%、或90%、或95%、或97%或99%相同,条件是已经从野生型序列中被修饰的核苷酸中的至少一个、或两个、或三个、或四个、或五个、或六个、或七个、或八个、或九个或十个或更多个、或全部没有从SEQ ID NO.:1中被修饰。在另一个实施方案中,分离的细胞进一步包含编码四环素激活蛋白的核酸和调节四环素激活蛋白表达的启动子,或基本上由其组成,或进一步由其组成。在一个实施方案中,调节四环素激活蛋白的表达的启动子是组成型启动子。在一个相关的实施方案中,该启动子是磷酸甘油酸激酶启动子(PGK)或CMV启动子。

[0128] 在一个特定的实施方案中,分离的细胞包含含有编码GALT蛋白或其生物等效物的SEQ ID NO.:1的多核苷酸的核酸,或基本上由其组成,或进一步由其组成。在一个相关的实施方案中,GALT的生物等效物包含在高严格条件下与SEQ ID NO.:1的互补序列杂交并编码GALT蛋白(例如,SEQ ID NO. 4)的核酸。在另一个实施方案中,其生物等效物包含与SEQ ID NO.:1具有至少80%序列同一性、或至少85%序列同一性、或至少90%序列同一性、或至少92%序列同一性、或至少95%序列同一性、或至少97%序列同一性、或至少98%序列同一性的核酸,并且在一方面,其中已经从野生型序列中被修饰的核苷酸中的至少一个、或两个、或三个、或四个、或五个、或六个、或七个、或八个、或九个、或十个或更多个、或全部没有从SEQ ID NO.:1中被修饰(参见图4)。在另一方面,GALT蛋白是野生型人GALT蛋白。

[0129] 本文所述的分离的细胞可以是以下物种的细胞中的任何一种:小鼠、大鼠、兔、猿猴、牛、绵羊、猪、犬、猫、农场动物、运动动物、宠物、马和灵长类动物、以及由其是人类细胞。

[0130] 在一些实施方案中,本文所述的分离的细胞包含一定水平的GALT蛋白。可以通过

选择产生所需蛋白质水平的合适的组成型启动子或通过使用调控蛋白质生成的量的诱导型系统来实现GALT蛋白质水平。这些启动子和诱导型系统已经在前面进行了描述。在一个实施方案中,分离的细胞包含至少约 $1 \times 10^{-7}$  ng、约 $3 \times 10^{-7}$  ng、约 $5 \times 10^{-7}$  ng、约 $7 \times 10^{-7}$  ng、约 $9 \times 10^{-7}$  ng、约 $1 \times 10^{-6}$  ng、约 $2 \times 10^{-6}$  ng、约 $3 \times 10^{-6}$  ng、约 $4 \times 10^{-6}$  ng、约 $6 \times 10^{-6}$  ng、约 $7 \times 10^{-6}$  ng、约 $8 \times 10^{-6}$  ng、约 $9 \times 10^{-6}$  ng、约 $10 \times 10^{-6}$  ng、约 $12 \times 10^{-6}$  ng、约 $14 \times 10^{-6}$  ng、约 $16 \times 10^{-6}$  ng、约 $18 \times 10^{-6}$  ng、约 $20 \times 10^{-6}$  ng、约 $25 \times 10^{-6}$  ng、约 $30 \times 10^{-6}$  ng、约 $35 \times 10^{-6}$  ng、约 $40 \times 10^{-6}$  ng、约 $45 \times 10^{-6}$  ng、约 $50 \times 10^{-6}$  ng、约 $55 \times 10^{-6}$  ng、约 $60 \times 10^{-6}$  ng、约 $65 \times 10^{-6}$  ng、约 $70 \times 10^{-6}$  ng、约 $75 \times 10^{-6}$  ng、约 $80 \times 10^{-6}$  ng、约 $85 \times 10^{-6}$  ng、约 $90 \times 10^{-6}$  ng、约 $95 \times 10^{-6}$  ng、约 $10 \times 10^{-5}$  ng、约 $20 \times 10^{-5}$  ng、约 $30 \times 10^{-5}$  ng、约 $40 \times 10^{-5}$  ng、约 $50 \times 10^{-5}$  ng、约 $60 \times 10^{-5}$  ng、约 $70 \times 10^{-5}$  ng、约 $80 \times 10^{-5}$  ng或约 $90 \times 10^{-5}$  ng的GALT蛋白,或基本上由其组成,或进一步由其组成。

#### [0131] 治疗组合物和方法

本公开提供了治疗半乳糖血症的治疗性基因载体组合物和疗法。半乳糖血症是一种罕见疾病,由GALT基因(半乳糖1-磷酸尿苷酰转移酶)中的隐性单基因缺陷引起。参见 [ghr.nlm.nih.gov/gene/GALT](http://ghr.nlm.nih.gov/gene/GALT)。在美国,约45,000个新生儿中有1个受其影响。幸运的是,在全国范围内进行的新生儿筛查程序可以及早发现它。唯一的治疗方法是从孩子的饮食中去除所有来源的乳糖和半乳糖。半乳糖限制的早期治疗在严重的疾病形式中是有益的,并且可以挽救生命,但是通常会存在长期并发症,例如语言和认知障碍,以及由于其他食物中含有的低水平的半乳糖和内源性生产的半乳糖引起的神经肌肉和卵巢毒性而导致的身体障碍。

[0132] 本文描述了载体的各种实施方案。在一个实施方案中,本文提供的载体和使用该载体的方法包含施用有效量的重组腺相关病毒(“AAV”)载体,或基本上由其组成,或进一步由其组成,所述重组腺相关病毒载体包含本公开提供的编码半乳糖-1-磷酸尿苷酰转移酶(GALT)的多核苷酸序列,或基本上由其组成,或进一步由其组成。在一方面,编码GALT的多核苷酸序列包含与SEQ ID NO:1至少85%、或90%、或95%、或97%或99%相同的核苷酸序列。在一方面,该序列与SEQ ID NO:1至少85%、或90%、或95%、或97%或99%相同,条件是已经从野生型序列中被修饰的核苷酸中的至少一个、或两个、或三个、或四个、或五个、或六个、或七个、或八个、或九个、或十个或更多个、或全部没有从SEQ ID NO:1中被修饰(参见图4)。在另一方面,编码GALT的多核苷酸序列包含如SEQ ID NO:1所示的核苷酸序列。在另一个实施方案中,多核苷酸序列编码SEQ ID NO:4的氨基酸序列。在一方面,多核苷酸和/或AAV载体含有报告基因荧光素酶或GFP,以确定载体的生物分布。在另一个实施方案中,多核苷酸和/或AAV载体包含多核苷酸序列,或基本上由其组成,或进一步由其组成,所述多核苷酸序列包含来自人(例如NCBI参考序列:NM\_000155.3或NM\_001258332.1)、小鼠(例如NCBI参考序列:NM\_001302511.1或NM\_016658.3)、大鼠(例如NCBI参考序列:NM\_001013089.2)、黑猩猩(例如NCBI参考序列:XM\_003951414.4或XM\_001163419.6)或其他物种的GALT基因序列。在另一个实施方案中,多核苷酸和/或AAV载体包含多核苷酸序列,或基本上由其组成,或进一步由其组成,所述多核苷酸序列包含人GALT基因的密码子优化版本(SEQ ID NO. 1和图4)。AAV载体中的一个被鉴定为pscAAV-CB-hGALT(SEQ ID NO.2)。在另一个实施方案中,AAV载体被鉴定为pAAV-CB-hGALT-WPREv2(SEQ ID NO.3)。pscAAV-CB-hGALT载体包含具有驱动人GALT

基因(半乳糖-1-磷酸尿苷酰转移酶)的密码子优化版本(SEQ ID NO. 1和图4)的小的SV40内含子的组成型CBA启动子。

[0133] 一方面,AAV载体包含编码人GALT蛋白(例如,SEQ ID NO.4)、小鼠GALT蛋白(例如,NCBI参考序列NP\_001289440.1)、大鼠GALT蛋白(例如,NCBI参考序列NP\_001013107.1)、黑猩猩GALT蛋白(例如,NCBI参考序列XP\_003951463.1)或来自其他物种的GALT蛋白的多核苷酸序列。在另一个实施方案中,AAV载体包含编码人GALT蛋白(例如,SEQ ID NO.4)的多核苷酸序列。

[0134] 一方面,AAV是AAV9血清型。替代血清型或经修饰的衣壳病毒可以被用于优化神经元趋向性。替代载体包括:用于比标准AAV9更高的神经元趋向性的经修饰的AAV9血清型载体,例如使用Cre-lox重组系统来鉴定神经靶向的载体的PHP.B。替代地,AAV PHP.B具有VP1的498位经修饰的氨基酸(从天冬酰胺到赖氨酸),以减少肝脏趋向性。突变多个氨基酸的AAVrh74的其他变体可以被用于包括脑在内的非常广泛的组织趋向性。

[0135] 在一方面,AAV载体被包含在经修饰的病毒衣壳蛋白内,所述病毒衣壳蛋白包含通过氨基酸置换或插入1至7个氨基酸来修饰的病毒衣壳蛋白,或基本上由其组成,或进一步由其组成。申请人已经生成了三个突变体。突变体之一(AAVmut4,在VP1衣壳的氨基酸502处的天冬酰胺变为异亮氨酸)增加对所有测试组织的全身基因递送,将转导效率提高高达56倍(根据组织,增加3至56倍)。另一个突变体(AAVmut5,在VP1衣壳的氨基酸505处色氨酸变为精氨酸)向心脏的基因递送比AAVrh74增加了几乎50倍。第三个突变体(AAVYIG591)靶向主要存在于被认为是肌肉干细胞的卫星细胞上的受体,尽管卫星细胞趋向性。值得注意的是,基于申请人对AAV晶体结构和 $\alpha 7\beta 1$ 整合蛋白的了解,设计AAVYIG591被认为对骨骼肌具有较高的亲和力且对肝脏具有较低的亲和力。

[0136] 不受理论的束缚,申请人期望通过增加到达肌肉的有效剂量而不增加对患者的总剂量来对患者实现治疗益处的方法。通过减少获得治疗益处所需的总剂量,更少的病毒抗原被递送给患者,理想地导致对载体的免疫应答降低和安全性提高。制造足够的基因治疗药物产品以进行后期临床试验是进一步发展的主要障碍。降低获得治疗益处的剂量需求将导致制造要求降低、制造成本降低、更快的临床试验进展以及更大的治疗更多患者的能力。

[0137] 因此,本公开涉及经修饰衣壳蛋白中包含的AAV载体、分离的多核苷酸、制备经修饰的衣壳蛋白的方法、重组病毒颗粒和用于产生经修饰的病毒颗粒的重组表达系统。本公开的一个方面涉及一种经修饰的病毒衣壳蛋白,其包含通过氨基酸置换或插入1至7个氨基酸来修饰的病毒衣壳蛋白,或基本上由其组成,或进一步由其组成。在一些实施方案中,病毒衣壳蛋白是VP1,任选地是AAV9、AAV PHP.B或AAVrh74。在进一步的实施方案中,修饰包含在AAVrh74的VP1的502位氨基酸处异亮氨酸取代天冬酰胺或等效修饰。在一些实施方案中,修饰包含在AAVrh74的VP1的氨基酸505处色氨酸被精氨酸取代。在一些实施方案中,修饰靶向主要存在于卫星细胞(任选地是肌肉干细胞)上的受体。在一些实施方案中,修饰是在AAVrh74的VP1的591位氨基酸处插入肽YIG。在一些实施方案中,所述肽对 $\alpha 7\beta 1$ 整合蛋白具有高亲和力 and/或位于可能改变正常rh74受体结合的区域中。

[0138] 本文公开了两个示例性载体的质粒骨架。第一载体含有CBA启动子/增强子-驱动由荧光素酶和增强的黄色荧光蛋白组成的报告融合蛋白的表达。在一方面,荧光素酶和增强的黄色荧光蛋白序列被删除。该载体可以作为单链病毒被包装在标准AAV9血清型衣壳或

突变衣壳内部。在另一方面, AAV包含具有驱动密码子优化的人GALT基因(SEQ ID NO 1)或其等效物的表达的CBA启动子/增强子的自互补载体。该载体可以被包装在标准AAV9血清型衣壳中。该载体可以提供高水平的人GALT蛋白。可以通过使用Xenogen IVIS(或类似物)在不同时间点进行体内荧光素酶染色,或在不同时间点收获组织并分析EYFP或荧光素酶活性的基因表达,从而评估luc-EYFP报道病毒的生物分布和基因表达。可以另外获得样品并通过qPCR评估载体基因组定量。施用可以开始于新生的-受影响的动物(出生后第1天或第2天)。在动物模型中,可以通过面部静脉注射老鼠幼畜(参见jove.com/video/52037/intravenous-injections-in-neonatal-mice),并在4周后通过GFP表达寻找载体的生物分布。由正常饮食喂养的母亲所产的纯合GALT缺陷小鼠可以很好地表明基因敲除(KO)小鼠模型中的载体趋向性和生物分布。

[0139] 本公开还提供了包含载体和经修饰的蛋白质、多核苷酸、载体、质粒、宿主细胞或表达系统中的一种或多种的组合物。本公开还提供了一种试剂盒,其包含经修饰的蛋白质、多核苷酸、载体、质粒、宿主细胞或表达系统中的一种或多种以及使用说明。可以将组合物配制成用于特定的施用方式。

#### [0140] 施用

在整个治疗过程中,本公开的重组多核苷酸、载体(例如, AAV)、病毒颗粒或组合物的施用可以以一个剂量连续或间歇地进行。施用可以通过任何合适的施用方式,包括但不限于:局部的、静脉内、动脉内、肌肉内、心内、鞘内、皮下、硬膜外、大脑内、脑室内、视网膜下、玻璃体内、关节内、眼内、腹膜内、子宫内、皮内、皮下、经皮、经粘膜和吸入。在一个实施方案中,重组载体或组合物通过肌肉内注射或静脉内注射被施用。在另一个实施方案中,全身性施用重组AAV载体或组合物。在另一个实施方案中,重组AAV载体或组合物通过注射、输注或植入进行肠胃外施用。在一方面, AAV载体或GALT多核苷酸被局部递送至肝脏。

[0141] 确定最有效的施用方式和剂量的方法是本领域技术人员已知的,并且将根据用于治疗的组合物、治疗的目的和所治疗的对象而变化。可以通过治疗医师选择剂量水平和方式进行单次或多次给药。值得注意的是,剂量会受到施用途径的影响。合适的剂型和施用试剂的方法是本领域已知的。每次施用的这种合适剂量的非限制性实施例可以低至 $1E+9$ 个载体基因组至多达 $1E+17$ 个载体基因组。

[0142] 在本文描述的方法的一些实施方案中,施用于对象的病毒颗粒(例如, AAV)的数量在约 $10^9$ 至约 $10^{17}$ 的范围内。在特定的实施方案中,向对象施用约 $10^{10}$ 至约 $10^{12}$ 、约 $10^{11}$ 至约 $10^{13}$ 、约 $10^{11}$ 至约 $10^{12}$ 、约 $10^{11}$ 至约 $10^{14}$ 、约 $5 \times 10^{11}$ 至约 $5 \times 10^{12}$ 、或约 $10^{12}$ 至约 $10^{13}$ 个病毒颗粒。

[0143] 在另一方面,本公开的多核苷酸、病毒颗粒和组合物可以与其他治疗(例如适合于半乳糖血症及其相关疾病或病症的那些被批准的治疗)组合施用。非限制性实施例包括用本公开的病毒载体或组合物治疗半乳糖血症,同时减少或消除对象饮食中的乳糖和/或半乳糖。

[0144] 当检测到以下一项或多项情况时,确定成功的治疗和/或修复:受治疗对象的疾病、病症或状况的一种或多种症状减轻或改善,对象的疾病、病症或状况的程度降低,稳定(即不恶化)疾病、病症或状况的状态,延缓或减缓疾病、病症或状况的进展,改善或减轻疾病、病症或状况。在一些实施方案中,通过检测从对象分离的一种或多种细胞、组织或器官中存在修复的靶多核苷酸来确定治疗的成功在一些实施方案中,通过检测由从对象分离的

一种或多种细胞、组织或器官中修复的靶多核苷酸编码的多肽的存在来确定治疗的成功。

[0145] 在一个实施方案中,重组多核苷酸和/或病毒载体可以修复对象中的GALT基因。

[0146] 在一些实施方案中,在成功治疗的细胞、组织、器官或对象中,修复的靶多核苷酸或多肽与未修复的靶多核苷酸或多肽的比例为约1.5:1、约2:1、约3:1、约4:1、约5:1、约6:1、约7:1、约8:1、约9:1、约10:1、约20:1、约50:1、约100:1、约1000:1、约10,000:1、约100,000:1或约1,000,000:1。修复的靶多核苷酸或多肽的量或比例可以通过本领域已知的任何方法来确定,包括但不限于Western印记杂交、Northern印记杂交、Southern印记杂交、PCR、测序、质谱、流式细胞术、免疫组织化学、免疫荧光、荧光原位杂交、下一代测序、免疫印迹和ELISA。

[0147] 试剂盒

在一些实施方案中,可以将本文所述的多核苷酸、试剂、载体或组合物组装成药物或诊断或研究试剂盒,以促进它们在治疗、诊断或研究应用中的用途。在一些实施方案中,本公开的试剂盒包括以下一种或多种:本文所述的经修饰的病毒衣壳蛋白、分离的多核苷酸、载体、宿主细胞、重组病毒颗粒、重组表达系统、经修饰的AAV、经修饰的细胞、分离的组织、组合物或药物组合物。

[0148] 在一些实施方案中,试剂盒进一步包括使用说明。具体而言,此类试剂盒可以包括一种或多种本文所述的试剂,以及描述这些试剂的预期应用和正确使用的说明。例如,在一个实施方案中,试剂盒可以包括用于混合试剂盒的一种或多种组分和/或分离和混合样品并施用于对象的说明。在某些实施方案中,试剂盒中的试剂为适合特定的应用和试剂施用方法的药物制剂和剂量。用于研究目的的试剂盒可以包含用于进行各种实验的适当浓度或数量的组分。

[0149] 试剂盒可以设计为促进本文所述方法的使用,并且可以采取多种形式。试剂盒的每种组合物,如果适用,可以以液体形式(例如溶液)或以固体形式(例如干粉)提供。在某些情况下,某些组合物可以是可组成的或可加工的(例如,制成活性形式),例如,通过添加试剂盒可以提供也可以不提供的合适的溶剂或其他种类的试剂(例如,水或细胞培养基)。在一些实施方案中,可以在保存溶液(例如,冷冻保存溶液)中提供组合物。保存溶液的非限制性实施例包括DMSO、低聚甲醛和CryoStor® (Stem Cell Technologies, Vancouver, Canada)。在一些实施方案中,保存溶液包含一定量的金属蛋白酶抑制剂。

[0150] 如本文所用,“说明/说明书”可以定义说明和/或促进的组成部分,并且通常涉及在所要求保护的方法、重组载体或组合物的包装上或与之相关的书面说明。说明还可以包括使得用户清楚地认识到说明与试剂盒相关联的以任何方式提供的任何口头或电子说明,例如视听(例如,录像带、DVD等)、互联网和/或基于Web的通信等。在一些实施方案中,书面说明采用对药品或生物产品的制造、使用或销售进行调控的政府机构规定的形式,该说明还可以反映出该机构对动物施用的生产、使用或销售的批准。

[0151] 在一些实施方案中,试剂盒在一个或多个容器中包含本文所述的任何一种或多种组分。因此,在一些实施方案中,试剂盒可以包括本文所述的容纳试剂的容器。所述试剂可以是液体、凝胶或固体(粉末)的形式。试剂可以无菌制备,包装在注射器中并冷藏运输。替代地,可以将其容纳在小瓶或其他容器中用于存储。第二容器可以具有无菌制备的其他试剂。替代地,试剂盒可以包括在注射器、小瓶、管或其他容器中预混合并运输的活性剂。试剂

盒可以具有向对象施用试剂所需的一种或多种或所有组分,例如注射器、局部施用装置或IV针管和袋。

[0152] 如本文所述的疗法可以与适当的诊断技术结合以识别和选择用于治疗的患者。例如,可以提供鉴定GALT基因突变的遗传测试。因此,具有突变的患者可以被鉴定为适合治疗。

## 实施例

### [0153] 体外施用

病毒颗粒是通过使用载体基因组的质粒、AAV rep和cap基因以及提供了生产AAV所需的腺病毒辅助功能的Ad辅助质粒的标准的三重转染方法制备的。使用PEI将这三个质粒转染到无血清悬浮生长的HEK293细胞中。转染后四天,通过标准方法收获和纯化病毒。如果基于相似的疾病模型和血清型静脉内给药,有效剂量范围为 $10^{12}$ 至 $10^{14}$  Vg/kg。AAVmut4包含对涉及肝脏特异性结合的区域修饰,从而使病毒对包括CNS和肌肉在内的许多其他目标组织具有更宽泛的趋向性。

[0154] 通过瞬时转染将质粒pAAV-CB-GALT-WPRE-kan(图1)转染到HEK293细胞中,两天后,裂解细胞,将收获的蛋白质在聚丙烯酰胺凝胶上电泳,转移到尼龙膜上,先用抗人GALT抗体探测,然后用IR染料标记的二抗探测。如图3所示,泳道1是分子量标记,其大小在Y轴上指示。泳道2是1  $\mu$ l经GALT转染的细胞裂解液,泳道3是0.1  $\mu$ l经GALT转染的细胞裂解液,泳道4是5  $\mu$ l用lucEYFP报告质粒转染的HEK293细胞裂解液,其用作转染和细胞对照,也显示出少量内源性表达的天然人类GALT蛋白,泳道5是空的,泳道6是HepG2细胞裂解物,显示天然人类GALT蛋白,泳道7是空的,泳道8-10显示浓度增加的细菌表达的GALT蛋白标准品,由于用于纯化的标签而具有稍高的迁移率。

### [0155] 体内施用

本文所公开的GALT基因疗法是高度创新的,因为首先,没有被批准用于治疗与经典半乳糖血症相关的长期并发症的AAV介导的基于基因的疗法或任何基于基因的疗法。其次,除了该技术之外,所公开的疗法还可以使用向肝脏的局部基因递送来使远处器官中异常的半乳糖代谢正常化。

[0156] 如Tang, M. et al. *Eur. J. Hum. Genet.*, (2014):1172-1179所述,使用了一种新的GALT基因陷阱(gene-trapped)(GalT缺陷)小鼠模型。使用LC-MS/MS方法(Li, Y. et al., *Mol. Genet. Metab.* (2011) 102(1):33-40),在纯合的GalT基因陷阱小鼠中证实了GALT活性的完全不存在。纯合GalT基因陷阱小鼠的进一步特征表明,新生的GalT缺陷幼畜中半乳糖敏感、成年雌性的生育力降低、两性的GalT缺陷小鼠的运动功能受损以及生长受限(Tang, et al. (2014), supra; Balakrishnan, B. et al. (2016) 470(1):205-212; Chen et al. (2017) *J. Inherit. Metab. Dis.* 40(1):131-137)。

[0157] 如图6所示,将AAV9-GALT载体(参见图1和2)转导至HEK293细胞系导致丰富的GALT蛋白的产生。可以通过将强CBA启动子/增强子替换为将会根据需要产生接近生理水平的GALT蛋白的启动子来修饰GALT蛋白的表达。

[0158] 生物分布:已经确定AAV9载体具有广泛的生物分布,包括中枢神经系统(CNS),而AAV8主要靶向小鼠的肝脏。因此,可以与AAV8载体一起施用GALT和密码子优化的GALT。通过

4周龄的野生型和GalT缺陷型小鼠的脉管系统递送载体可以评估两组之间生物分布的任何差异。首先,通过尾静脉向雄性和雌性小鼠(每组每个性别N=5,N基于FDA的毒性/生物分布研究指南,该指南列于标题为“行业指南:基因治疗临床试验-观察对象的延迟不良事件”的文件中)施用 $1E+12$  Vg的表达lucEYFP报告基因构建体的每个AAV载体。在处死前,每周对小鼠成像,进行4周。对照是注射无GALT基因插入的载体的年龄和性别相匹配的动物。4周后,处死动物并用固定剂灌注,并分离出脑、心脏、肾脏、肝脏、眼睛和性腺。将组织切成薄片,以进行EYFP显微镜染色,而每个组织的一部分将用于qPCR测定每微克基因组DNA的载体基因组。在另一个对照施用中,用饮食中10%和20%的半乳糖攻击小鼠。

[0159] 剂量研究:分别给雄性和雌性4周龄GalT缺陷小鼠静脉内注射三种不同剂量的AAV9-GALT(或AAV8-GALT)载体之一,每只小鼠分别为 $1E+10$ 、 $1E+11$ 或 $1E+12$  vg,每组12只小鼠(每组每个性别N=6),再加上所选血清型的空载体注射对照组。监测动物长达6个月,并检查运动协调和生殖健康状况的潜在改善。还监视和分析动物的治疗的任何明显的不利生理作用,例如体重减轻或不活动。

[0160] 行为分析:测试小鼠(每组n=12)在6个月大时的行为表现,以量化治疗对神经系统疾病的功能影响。行为测试包括认知和Morris水迷宫中的游泳能力,以及旋转脚架上的运动功能。为了确定小鼠的与共济失调相关的运动障碍,将使用一种改良的旋转脚架,以鼓励动物沿着杆在分枝之间交替行走,而不是被动地使用其身体作为支撑。为了进行培训,在两分钟内将速度逐渐提高到6 RPM。休息两分钟后,以6 RPM的速度测试小鼠,最多两分钟(如果动物未通过测试,则时间更短)。对于每只小鼠,可以以12 RPM重复进行该试验,以提高与野生型(WT)正常小鼠相比的攻击强度。在五分钟的休息时间后,以各种速度对每只小鼠进行三次测试。在6个月末,将所有小鼠实施安乐死,并通过Western评估GALT基因表达,通过qPCR评估脑和其他组织的各个区域中载体基因组含量,以及通过董事会认证的病理学家进行用于病理学的H&E组织学检查。

[0161] 雌性小鼠的生育力评估:通过在固定的时间间隔监测雌性小鼠的发情周期并在六个月末进行卵泡计数来评估雌性小鼠的生殖健康。单独的发情周期和卵泡计数正常化的任何迹象都是改善这些动物生育力的良好指标。

[0162] 等效物

应当理解,尽管已经结合以上实施方案描述了本发明,但是前述说明和实施例旨在说明而不是限制本发明的范围。在本发明范围内的其他方面,优点和修改对于本发明所属领域的技术人员将是显而易见的。

[0163] 另此外,在以马库什组描述本发明的特征或方面的情况中,本领域技术人员将认识到,还借此以该马库什组中的任何单独成员或成员的亚组描述了本发明。

[0164] 在本文中提及的所有技术和专利出版物均通过引用并入本文。

[0165] 参考文献

1. Fang, H., et al., Comparison of adeno-associated virus serotypes and delivery methods for cardiac gene transfer. *Hum Gene Ther Methods*, 2012. 23(4): p. 234-41.

2. Bianconi, E., et al., An estimation of the number of cells in the human body. *Ann Hum Biol*, 2013. 40(6): p. 463-71.

3. Tran, T., et al., Laminin drives survival signals to promote a contractile smooth muscle phenotype and airway hyperreactivity. *FASEB J*, 2013. 27(10): p. 3991-4003.

4. Mori, S., et al., Biodistribution of a low dose of intravenously administered AAV-2, 10, and 11 vectors to cynomolgus monkeys. *Jpn J Infect Dis*, 2006. 59(5): p. 285-93.

5. Grimm, D., et al., In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. *J Virol*, 2008. 82(12): p. 5887-911.

6. Asokan, A., et al., Reengineering a receptor footprint of adeno-associated virus enables selective and systemic gene transfer to muscle. *Nat Biotechnol*, 2010. 28(1): p. 79-82.

7. Pulicherla, N., et al., Engineering liver-detargeted AAV9 vectors for cardiac and musculoskeletal gene transfer. *Mol Ther*, 2011. 19(6): p. 1070-8.

8. DiMattia, M.A., et al., Structural insight into the unique properties of adeno-associated virus serotype 9. *J Virol*, 2012. 86(12): p. 6947-58.

9. Li, C., et al., Single amino acid modification of adeno-associated virus capsid changes transduction and humoral immune profiles. *J Virol*, 2012. 86(15): p. 7752-9.

10. Raupp, C., et al., The threefold protrusions of adeno-associated virus type 8 are involved in cell surface targeting as well as postattachment processing. *J Virol*, 2012. 86(17): p. 9396-408.

11. Govindasamy, L., et al., Structural insights into adeno-associated virus serotype 5. *J Virol*, 2013. 87(20): p. 11187-99.

12. Bentzinger, C.F., et al., Cellular dynamics in the muscle satellite cell niche. *EMBO Rep*, 2013. 14(12): p. 1062-72.

13. Wang, Y.X., N.A. Dumont, and M.A. Rudnicki, Muscle stem cells at a glance. *J Cell Sci*, 2014. 127(21): p. 4543-8.

14. Loiler, S.A., et al., Targeting recombinant adeno-associated virus vectors to enhance gene transfer to pancreatic islets and liver. *Gene Ther*, 2003. 10(18): p. 1551-8.

[0166] SEQ ID NO 1

密码子优化的GALT序列

```
1 ATGAGCAGAA GCGGCACCGA CCCTCAGCAG AGACAGCAGG CCTCTGAAGC CGATGCCGCC
61 GCTGCCACCT TCAGAGCCAA TGACCACCAG CACATCCGGT ACAACCCCTT GCAGGACGAG
121 TGGGTGCTGG TGTCCGCCA CAGAATGAAG AGGCCTTGGC AGGGCCAGGT GGAACCCAG
181 CTGCTGAAAA CCGTGCCAG ACACGACCCC CTGAACCCCTC TGTGTCTCTGG CGCCATTAGA
```

241 GCCAACGGCG AAGTGAACCC CCAGTACGAC AGCACCTTCC TGTTTCGACAA CGACTTCCCC  
 301 GCCCTGCAGC CTGATGCCCC ATCTCCTGGA CCTAGCGACC ACCCTCTGTT CCAGGCCAAG  
 361 TCTGCCAGAG GCGTGTGCAA AGTGATGTGC TTCCACCCTT GGAGCGACGT GACCCTGCCC  
 421 CTGATGAGCG TGCCAGAGAT CAGAGCCGTG GTGGATGCCT GGGCCAGCGT GACAGAAGAA  
 481 CTGGGAGCCC AGTACCCTG GGTGCAGATC TTCGAGAACA AGGGCGCCAT GATGGGCTGC  
 541 AGCAACCCCC ACCCTCACTG TCAAGTGTGG GCCAGCAGCT TCCTGCCCGA TATCGCCCAG  
 601 CGGGAAGAGA GAAGCCAGCA GGCTTACAAG AGCCAGCAGC GCGAGCCCCT GCTGATGGAA  
 661 TACTCCAGAC AGGAACTGCT GCGGAAAGAA CGGCTGGTGC TGACCAGCGA GCACTGGCTG  
 721 GTGCTGGTGC CTTTTTGGGC CACATGGCCC TACCAGACCC TGCTGCTGCC TAGAAGGCAC  
 781 GTGCGGAGAC TGCCTGAGCT GACACCCGCC GAGAGAGATG ACCTGGCCAG CATCATGAAG  
 841 AAAGTCTGA CCAAATACGA CAACCTGTTC GAGACCAGCT TCCCCTACAG CATGGGCTGG  
 901 CACGGCGCTC CTACAGGATC TGAGGCTGGC GCCAACTGGA ACCACTGGCA GCTGCACGCC  
 961 CACTACTACC CCCCACTGCT GAGATCTGCC ACCGTGCGGA AGTTCATGGT GGGATACGAG  
 1021 ATGCTGGCTC AGGCCAGAG AGATCTGACC CCTGAACAGG CCGCCGAACG

GCTGAGAGCA

1081 CTGCCCGAAG TGCCTACCA CCTGGGACAG AAGGACAGAG AGACAGCCAC  
AATCGCCTGA。

[0167] SEQ ID NO 2

突变的ITR:核苷酸(NT):1-106

CMV增强子:NT:153-432

鸡β-肌动蛋白启动子:NT 439-708

经修饰的SV40内含子:774-833

人GALT:1004-2143

牛生长激素多腺苷酸化信号:NT 2186-2332

ITR:NT:2412-2552

氨苄青霉素抗性基因:NT 3455-4315

质粒复制起点(ori):NT:4470-5089

1 CTGCGCGCTC GCTCGCTCAC TGAGGCCGCC CGGGCAAAGC CCGGGCGTCG GGCGACCTTT  
 61 GGTCGCCCGG CCTCAGTGAG CGAGCGAGCG CGCAGAGAGG GAGTGAATT CACGCGTGGA  
 121 TCTGAATTCA ATTCACGCGT GGTACCTCTG GTCGTTACAT AACTTACGGT AAATGGCCCG  
 181 CCTGGCTGAC CGCCCAACGA CCCCCGCCA TTGACGTCAA TAATGACGTA TGTTCCATA  
 241 GTAACGCCAA TAGGGACTTT CCATTGACGT CAATGGGTGG AGTATTTACG GTAAACTGCC  
 301 CACTTGGCAG TACATCAAGT GTATCATATG CCAAGTACGC CCCCTATTGA CGTCAATGAC  
 361 GGTAATGGC CCGCCTGGCA TTATGCCAG TACATGACCT TATGGGACTT TCCTACTTGG  
 421 CAGTACATCT ACTCGAGGCC ACGTTCTGCT TCACTCTCCC CATCTCCCC CCCTCCCCAC  
 481 CCCCAATTTT GTATTTATTT ATTTTTTAAT TATTTTGTGC AGCGATGGGG GCGGGGGGGG  
 541 GGGGGGGGCG CGCGCCAGGC GGGGCGGGG GGGGCGAGGG GCGGGGCGGG GCGAGGCGGA  
 601 GAGGTGCGGC GGCAGCCAAT CAGAGCGGCG CGCTCCGAAA GTTTCCTTTT ATGGCGAGGC  
 661 GCGGGCGGCG GCGGCCCTAT AAAAAGCGAA GCGCGCGGCG GCGGGGAGCG GGATCAGCCA

721 CCGCGGTGGC GGCCTAGAGT CGACGAGGAA CTGAAAAACC AGAAAGTTAA CTGGTAAGTT  
781 TAGTCTTTTT GTCTTTTATT TCAGGTCCCG GATCCGGTGG TGGTGCAAAT CAAAGAACTG  
841 CTCCTCAGTG GATGTTGCCT TTA CTCTAG GCCTGTACGG AAGTGTTACT TCTGCTCTAA  
901 AAGCTGCGGA ATTGTACCCG CGGCCGATCC ACCGGTCTTA AGGGCCGAGG CGGCCAGATC  
961 TTTCGAAGAT ATCGGCGCCG CTAGCGCGGC CGCAGCTGCC ACCATGAGCA GAAGCGGCAC  
1021 CGACCCTCAG CAGAGACAGC AGGCCTCTGA AGCCGATGCC GCCGCTGCCA  
CCTTCAGAGC  
1081 CAATGACCAC CAGCACATCC GGTACAACCC CCTGCAGGAC GAGTGGGTGC  
TGGTGTCCGC  
1141 CCACAGAATG AAGAGGCCTT GGCAGGGCCA GGTGGAACCC CAGCTGCTGA  
AAACCGTGCC  
1201 CAGACACGAC CCCCTGAACC CTCTGTGTCC TGGCGCCATT AGAGCCAACG  
GCGAAGTGAA  
1261 CCCCCAGTAC GACAGCACCT TCCTGTTCGA CAACGACTTC CCCGCCCTGC  
AGCCTGATGC  
1321 CCCATCTCCT GGACCTAGCG ACCACCCTCT GTTCCAGGCC AAGTCTGCCA  
GAGGCGTGTG  
1381 CAAAGTGATG TGCTTCCACC CTTGGAGCGA CGTGACCCTG CCCCTGATGA  
GCGTGCCAGA  
1441 GATCAGAGCC GTGGTGGATG CCTGGGCCAG CGTGACAGAA GAACTGGGAG  
CCCAGTACCC  
1501 CTGGGTGCAG ATCTTCGAGA ACAAGGGCGC CATGATGGGC TGCAGCAACC  
CCCACCCTCA  
1561 CTGTCAAGTG TGGGCCAGCA GCTTCCTGCC CGATATCGCC CAGCGGGAAG  
AGAGAAGCCA  
1621 GCAGGCTTAC AAGAGCCAGC ACGGCGAGCC CCTGCTGATG GAATACTCCA  
GACAGGAACT  
1681 GCTGCGGAAA GAACGGCTGG TGCTGACCAG CGAGCACTGG CTGGTGCTGG  
TGCCTTTTTG  
1741 GGCCACATGG CCCTACCAGA CCCTGCTGCT GCCTAGAAGG CACGTGCGGA  
GACTGCCTGA  
1801 GCTGACACCC GCCGAGAGAG ATGACCTGGC CAGCATCATG AAGAACTGC  
TGACCAAATA  
1861 CGACAACCTG TTCGAGACCA GCTTCCCCTA CAGCATGGGC TGGCACGGCG  
CTCCTACAGG  
1921 ATCTGAGGCT GGCGCCAACT GGAACCACTG GCAGCTGCAC GCCCACTACT  
ACCCCCACT  
1981 GCTGAGATCT GCCACCGTGC GGAAGTTCAT GGTGGGATAC GAGATGCTGG  
CTCAGGCCCA

2041 GAGAGATCTG ACCCCTGAAC AGGCCGCCGA ACGGCTGAGA GCACTGCCCC  
AAGTGCACTA  
2101 CCACCTGGGA CAGAAGGACA GAGAGACAGC CACAATCGCC TGAAGTCAAG  
CTTATCGATA  
2161 CCGTCGACTA GAGCTCGCTG ATCAGCCTCG ACTGTGCCTT CTAGTTGCCA  
GCCATCTGTT  
2221 GTTTGCCCCT CCCCCGTGCC TTCCTTGACC CTGGAAGGTG CCACTCCCAC  
TGTCCTTTCC  
2281 TAATAAAATG AGGAAATTGC ATCGCATTGT CTGAGTAGGT GTCATTCTAT  
TCTGGGGGGT  
2341 GGGGTGGGGC AGGACAGCAA GGGGGAGGAT TGGGAAGACA ATAGCAGGCA  
TGCTGGGGAG  
2401 AGATCGATCT GAGGAACCCC TAGTGATGGA GTTGGCCACT CCCTCTCTGC  
GCGCTCGCTC  
2461 GCTCACTGAG GCCGGGCGAC CAAAGGTCGC CCGACGCCCG GGCTTTGCC  
GGCGGCCTC  
2521 AGTGAGCGAG CGAGCGCGCA GAGAGGGAGT GGCCCCCCCC CCCCCCCCC  
CGGCGATTCT  
2581 CTTGTTTGCT CCAGACTCTC AGGCAATGAC CTGATAGCCT TTGTAGAGAC  
CTCTCAAAAA  
2641 TAGCTACCCT CTCCGGCATG AATTTATCAG CTAGAACGGT TGAATATCAT  
ATTGATGGTG  
2701 ATTTGACTGT CTCCGGCCTT TCTCACCCGT TTGAATCTTT ACCTACACAT  
TACTCAGGCA  
2761 TTGCATTTAA AATATATGAG GGTTCATAAAA ATTTTTATCC TTGCGTTGAA  
ATAAAGGCTT  
2821 CTCCCGCAA AGTATTACAG GGCATAATG TTTTGGTAC AACCGATTTA  
GCTTTATGCT  
2881 CTGAGGCTTT ATTGCTTAAT TTTGCTAATT CTTTGCCTTG CCTGTATGAT  
TTATTGGATG  
2941 TTGGAATCGC CTGATGCGGT ATTTTCTCCT TACGCATCTG TCGGGTATTT  
CACACCGCAT  
3001 ATGGTGCACT CTCAGTACAA TCTGCTCTGA TGCCGCATAG TTAAGCCAGC  
CCCACACCC  
3061 GCCAACACTA TGGTGCACTC TCAGTACAAT CTGCTCTGAT GCCGCATAGT  
TAAGCCAGCC  
3121 CCGACACCCG CCAACACCCG CTGACGCGCC CTGACGGGCT TGTCTGCTCC  
CGGCATCCGC  
3181 TTACAGACAA GCTGTGACCG TCTCCGGGAG CTGCATGTGT CAGAGGTTTT

CACCGTCATC

3241 ACCGAAACGC GCGAGACGAA AGGGCCTCGT GATACGCCTA TTTTATAGG  
TTAATGTCAT

3301 GATAATAATG GTTTCTTAGA CGTCAGGTGG CACTTTTCGG GGAAATGTGC  
GCGGAACCCC

3361 TATTTGTTTA TTTTCTAAA TACATTCAAA TATGTATCCG CTCATGAGAC  
AATAACCCTG

3421 ATAAATGCTT CAATAATATT GAAAAAGGAA GAGTATGAGT ATTCAACATT  
TCCGTGTCGC

3481 CCTTATTCCC TTTTTGCGG CATTTTGCCT TCCTGTTTTT GCTCACCCAG  
AAACGCTGGT

3541 GAAAGTAAAA GATGCTGAAG ATCAGTTGGG TGCACGAGTG GGTTACATCG  
AACTGGATCT

3601 CAACAGCGGT AAGATCCTTG AGAGTTTTCG CCCC GAAGAA CGTTTTCCAA  
TGATGAGCAC

3661 TTTTAAAGTT CTGCTATGTG GCGCGGTATT ATCCCGTATT GACGCCGGGC  
AAGAGCAACT

3721 CGGTCGCCGC ATACACTATT CTCAGAATGA CTTGGTTGAG TACTCACCAG  
TCACAGAAAA

3781 GCATCTTACG GATGGCATGA CAGTAAGAGA ATTATGCAGT GCTGCCATAA  
CCATGAGTGA

3841 TAACACTGCG GCCAACTTAC TTCTGACAAC GATCGGAGGA CCGAAGGAGC  
TAACCGCTTT

3901 TTTGCACAAC ATGGGGGATC ATGTAACTCG CCTTGATCGT TGGGAACCGG  
AGCTGAATGA

3961 AGCCATACCA AACGACGAGC GTGACACCAC GATGCCTGTA GCAATGGCAA  
CAACGTTGCG

4021 CAAACTATTA ACTGGCGAAC TACTTACTCT AGCTTCCCGG CAACAATTAA  
TAGACTGGAT

4081 GGAGGCGGAT AAAGTTGCAG GACCACTTCT GCGCTCGGCC CTTCCGGCTG  
GCTGGTTTAT

4141 TGCTGATAAA TCTGGAGCCG GTGAGCGTGG GTCTCGCGGT ATCATTGCAG  
CACTGGGGCC

4201 AGATGGTAAG CCCTCCCGTA TCGTAGTTAT CTACACGACG GGGAGTCAGG  
CAACTATGGA

4261 TGAACGAAAT AGACAGATCG CTGAGATAGG TGCCTCACTG ATTAAGCATT  
GGTAACTGTC

4321 AGACCAAGTT TACTCATATA TACTTTAGAT TGATTTAAAA CTCATTTTT  
AATTTAAAAG

4381 GATCTAGGTG AAGATCCTTT TTGATAATCT CATGACCAAA ATCCCTTAAC  
GTGAGTTTTC

4441 GTTCCACTGA GCGTCAGACC CCGTAGAAAA GATCAAAGGA TCTTCTTGAG  
ATCCTTTTTT

4501 TCTGCGCGTA ATCTGCTGCT TGCAAACAAA AAAACCACCG CTACCAGCGG  
TGGTTTGTTT

4561 GCCGGATCAA GAGCTACCAA CTCTTTTTTCC GAAGGTA ACT GGCTTCAGCA  
GAGCGCAGAT

4621 ACCAAATACT GTTCTTCTAG TGTAGCCGTA GTTAGGCCAC CACTTCAAGA  
ACTCTGTAGC

4681 ACCGCCTACA TACCTCGCTC TGCTAATCCT GTTACCAGTG GCTGCTGCCA  
GTGGCGATAA

4741 GTCGTGTCTT ACCGGGTTGG ACTCAAGACG ATAGTTACCG GATAAGGCGC  
AGCGGTCGGG

4801 CTGAACGGGG GGTTCGTGCA CACAGCCCAG CTTGGAGCGA ACGACCTACA  
CCGAACTGAG

4861 ATACCTACAG CGTGAGCTAT GAGAAAGCGC CACGCTTCCC GAAGGGAGAA  
AGGCGGACAG

4921 GTATCCGGTA AGCGGCAGGG TCGGAACAGG AGAGCGCACG AGGGAGCTTC  
CAGGGGAAA

4981 CGCCTGGTAT CTTTATAGTC CTGTCGGGTT TCGCCACCTC TGA CTTGAGC  
GTCGATTTTT

5041 GTGATGCTCG TCAGGGGGGC GGAGCCTATG GAAAAACGCC AGCAACGCGG  
CCTTTTTACG

5101 GTTCCTGGCC TTTTGCTGGC CTTTGCTCA CATGTTCTTT CCTGCGTTAT  
CCCCTGATTC

5161 TGTGGATAAC CGTATTACCG CCTTTGAGTG AGCTGATACC GCTCGCCGCA  
GCCGAACGAC

5221 CGAGCGCAGC GAGTCAGTGA GCGAGGAAGC GGAAGAGCGC CCAATACGCA  
AACCGCTCT

5281 CCCC GCGCGT TGGCCGATTC ATTAATGCAG CTGGCGTAAT AGCGAAGAGG  
CCCGCACCGA

5341 TCGCCCTTCC CAACAGTTGC GCAGCCTGAA TGGCGAATGG CGATTCCGTT  
GCAATGGCTG

5401 GCGGTAATAT TGTTCTGGAT ATTACCAGCA AGGCCGATAG TTTGAGTTCT  
TCTACTCAGG

5461 CAAGTGATGT TATTACTAAT CAAAGAAGTA TTGCGACAAC GGTTAATTTG  
CGTGATGGAC

5521 AGACTCTTTT ACTCGGTGGC CTC ACTGATT ATAAAAACAC TTCTCAGGAT

TCTGGCGTAC

5581 CGTTCCTGTC TAAAATCCCT TTAATCGGCC TCCTGTTTAG CTCCCGCTCT  
GATTCTAACG

5641 AGGAAAGCAC GTTATACGTG CTCGTCAAAG CAACCATAGT ACGCGCCCTG  
TAGCGGCGCA

5701 TTAAGCGCGG CGGGTGTGGT GGTTACGCGC AGCGTGACCG CTACACTTGC  
CAGCGCCCTA

5761 GCGCCCCTC CTTTCGCTTT CTCCCTTCC TTTCTCGCCA CGTTCGCCGG  
CTTTCCCGT

5821 CAAGCTCTAA ATCGGGGGCT CCCTTTAGGG TTCCGATTTA GTGCTTTACG  
GCACCTCGAC

5881 CCCAAAAAAC TTGATTAGGG TGATGGTTCA CGTAGTGGGC CATCGCCCTG  
ATAGACGGTT

5941 TTTCGCCCTT TGACGTTGGA GTCCACGTC TTTAATAGTG GACTCTTGTT  
CCAAACTGGA

6001 ACAACACTCA ACCCTATCTC GGTCTATTCT TTTGATTTAT AAGGGATTTT  
GCCGATTTTCG

6061 GCCTATTGGT TAAAAAATGA GCTGATTTAA CAAAAATTTA ACGCGAATTT  
TAACAAAATA

6121 TTAACGCTTA CAATTTAAAT ATTTGCTTAT ACAATCTTCC TGTTTTTGGG  
GCTTTTCTGA

6181 TTATCAACCG GGGTACATAT GATTGACATG CTAGTTTTAC GATTACCGTT CATCGCC.

[0168] SEQ ID NO 3

突变的ITR:核苷酸(NT):1-145

CMV增强子:NT:190-551

鸡β-肌动蛋白启动子:NT 552-834

嵌合内含子:929-1103

人GALT:1160-2299

土拨鼠肝炎病毒Pos反应元件(WPRE):NT 2334-2927

Poly A信号(BGHPA):NT 2942-3176

ITR:NT:3449-3592

KAN基因:NT 5145-5960

1 GCCAATACG CAAACCGCCT CTCCCCGCGC GTTGCCGAT TCATTAATGC AGCTGGCGCG

61 CTCGCTCGCT CACTGAGGCC GCCCGGGCAA AGCCCGGGCG TCGGGCGACC TTTGGTCGCC

121 CGGCCTCAGT GAGCGAGCGA GCGCGCAGAG AGGGAGTGGC CAACTCCATC ACTAGGGGTT

181 CTTGTAGTT AATGATTAAC CCGCCATGCT AATTATCTAC GTAGCCATGT CTAGACAGCC

241 ACTATGGGTC TAGGCTGCCC ATGTAAGGAG GCAAGGCCTA GTTATTAATA GTAATCAATT

301 ACGGGTCAT TAGTTCATAG CCCATATATG GAGTTCCGCG TTACATAACT TACGGTAAAT

361 GGCCCGCCTG GCTGACCGCC CAACGACCCC CGCCATTGA CGTCAATAAT GACGTATGTT

421 CCCATAGTAA CGCCAATAGG GACTTTCCAT TGACGTCAAT GGGTGGACTA TTTACGGTAA  
481 ACTGCCCACT TGGCAGTACA TCAAGTGTAT CATATGCCAA GTACGCCCCC TATTGACGTC  
541 AATGACGGTA AATGGCCCGC CTGGCATTAT GCCCAGTACA TGACCTTATG GGACTTTCCT  
601 ACTTGGCAGT ACATCTACGT ATTAGTCATC GCTATTACCA TGGTCGAGGT GAGCCCCACG  
661 TTCTGCTTCA CTCTCCCAT CTCCCCCCC TCCCCACCC CAATTTTGTA TTTATTTATT  
721 TTTAATTAT TTTGTGCAGC GATGGGGCG GGGGGGGGG GGGGGCGGC GCCAGGCGGG  
781 GCGGGCGGG GCGAGGGCG GGGCGGGCG AGGCGGAGAG GTGCGGCGG AGCCAATCAG  
841 AGCGGCGGC TCCGAAAGTT TCCTTTTATG GCGAGGCGG GCGGCGGC GCCCTATAAA  
901 AAGCGAAGCG CGCGGCGGG GGGAGTCGCT GCGACGCTGC CTTGCCCGG TGCCCCGCTC  
961 CGCCGCCGCC TCGCGCCGCC CGCCCCGGCT CTGACTGACC GCGTTACTCC CACAGGTGAG  
1021 CGGGCGGGAC GGCCCTTCTC CTCCGGGCTG TAATTAGCGC TTGGTTTAAAT  
GACGGCTTGT  
1081 TTCTTTTCTG TGGCTGCGTG AAAGCCTTGA GGGGCTCCGG GAGCTAGAGC  
CTCTGCTAAC  
1141 CATGTTTCATG CCTTCTTCTT TTCCTACAG CTCCTGGGCA ACGTGCTGGT  
TATTGTGCTG  
1201 TCTCATCATT TTGGCAAAGA ATTCTAGCGC GGCCGCAGCT GCCACCATGA  
GCAGAAGCGG  
1261 CACCGACCCT CAGCAGAGAC AGCAGGCCTC TGAAGCCGAT GCCGCCGCTG  
CCACCTTCAG  
1321 AGCCAATGAC CACCAGCACA TCCGGTACAA CCCCTGCAG GACGAGTGGG  
TGCTGGTGTC  
1381 CGCCCACAGA ATGAAGAGGC CTTGGCAGGG CCAGGTGGAA CCCAGCTGC  
TGAAAACCGT  
1441 GCCCAGACAC GACCCCCTGA ACCCTCTGTG TCCTGGCGCC ATTAGAGCCA  
ACGGCGAAGT  
1501 GAACCCCCAG TACGACAGCA CCTTCTGTT CGACAACGAC TTCCCCGCC  
TGCAGCCTGA  
1561 TGCCCCATCT CCTGGACCTA GCGACCACC TCTGTTCCAG GCCAAGTCTG  
CCAGAGGCGT  
1621 GTGCAAAGTG ATGTGCTTCC ACCCTTGGAG CGACGTGACC CTGCCCTGA  
TGAGCGTGCC  
1681 AGAGATCAGA GCCGTGGTGG ATGCCTGGGC CAGCGTGACA GAAGAACTGG  
GAGCCCAGTA  
1741 CCCCTGGGTG CAGATCTTCG AGAACAAGGG CGCCATGATG GGCTGCAGCA  
ACCCCCACCC  
1801 TCACTGTCAA GTGTGGGCCA GCAGCTTCTT GCCCGATATC GCCCAGCGGG  
AAGAGAGAAG  
1861 CCAGCAGGCT TACAAGAGCC AGCACGGCGA GCCCCTGCTG ATGGAATACT

CCAGACAGGA

1921 ACTGCTGCGG AAAGAACGGC TGGTGTGAC CAGCGAGCAC TGGCTGGTGC  
TGGTGCCTTT

1981 TTGGGCCACA TGGCCCTACC AGACCCTGCT GCTGCCTAGA AGGCACGTGC  
GGAGACTGCC

2041 TGAGCTGACA CCCGCCGAGA GAGATGACCT GGCCAGCATC ATGAAGAAAC  
TGCTGACCAA

2101 ATACGACAAC CTGTTCGAGA CCAGCTTCCC CTACAGCATG GGCTGGCAGC  
GCGCTCCTAC

2161 AGGATCTGAG GCTGGCGCCA ACTGGAACCA CTGGCAGCTG CACGCCCACT  
ACTACCCCC

2221 ACTGCTGAGA TCTGCCACCG TCGGAAGTT CATGGTGGGA TACGAGATGC  
TGGCTCAGGC

2281 CCAGAGAGAT CTGACCCCTG AACAGGCCGC CGAACGGCTG AGAGCACTGC  
CCGAAGTGCA

2341 CTACCACCTG GGACAGAAGG ACAGAGAGAC AGCCACAATC GCCTGAAGTC  
AAGCTTATCG

2401 ATAATCAACC TCTGGATTAC AAAATTTGTG AAAGATTGAC TGGTATTCTT  
AACTATGTTG

2461 CTCCTTTTAC GCTATGTGGA TACGCTGCTT TAATGCCTTT GTATCATGCT  
ATTGCTTCCC

2521 GTATGGCTTT CATTTTCTCC TCCTTGTATA AATCCTGGTT GCTGTCTCTT  
TATGAGGAGT

2581 TGTGGCCCGT TGTCAGGCAA CGTGGCGTGG TGTGCACTGT GTTTGCTGAC  
GCAACCCCA

2641 CTGGTTGGGG CATTGCCACC ACCTGTCAGC TCCTTTCCGG GACTTTCGCT  
TTCCCCCTCC

2701 CTATTGCCAC GCGGAACTC ATCGCCGCCT GCCTTGCCCG CTGCTGGACA  
GGGGCTCGGC

2761 TGTTGGGCAC TGACAATTCC GTGGTGTGTG CGGGGAAATC ATCGTCCTTT  
CCTTGCTGC

2821 TCGCCTGTGT TGCCACCTGG ATTCTGCGCG GGACGTCCTT CTGCTACGTC  
CCTTCGCCCC

2881 TCAATCCAGC GGACCTTCCT TCCCGCGGCC TGCTGCCGGC TCTGCGGCCT  
CTTCCGCGTC

2941 TTCGCCTTCG CCCTCAGACG AGTCGGATCT CCCTTTGGGC CGCCTCCCCG  
CATCGATAAC

3001 GTCGAGGCCG CAATAAAAGA TCTTTATTTT CATTAGATCT GTGTGTTGGT  
TTTTTGTGTG

3061 TCTAGACATG GCTACGTAGA TAATTAGCAT GGCGGGTTAA TCATTAACTA  
CAAGGAACCC

3121 CTAGTGATGG AGTTGGCCAC TCCCTCTCTG CGCGCTCGCT CGCTCACTGA  
GGCCGGGCGA

3181 CCAAAGGTCG CCCGACGCCC GGGCTTTGCC CGGGCGGCCT CAGTGAGCGA  
GCGAGCGCGC

3241 CAGCTGGCGT AATAGCGAAG AGGCCCGCAC CGATCGCCCT TCCCAACAGT  
TGCGCAGCCT

3301 GAATGGCGAA TGGAAGTTCC GTTGCAATGG CTGGCGGTAA TATTGTTCTG  
GATATTACCA

3361 GCAAGGCCGA TAGTTTGAGT TCTTCTACTC AGGCAAGTGA TGTTATTACT  
AATCAAAGAA

3421 GTATTGCGAC AACGGTTAAT TTGCGTGATG GACAGACTCT TTTACTCGGT  
GGCCTCACTG

3481 ATTATAAAAA CACTTCTCAG GATTCTGGCG TACCGTTCCT GTCTAAAATC  
CCTTTAATCG

3541 GCCTCCTGTT TAGCTCCCGC TCTGATTCTA ACGAGGAAAG CACGTTATAC  
GTGCTCGTCA

3601 AAGCAACCAT AGTACGCGCC CTGTAGCGGC GCATTAAGCG CGGCGGGTGT  
GGTGGTTACG

3661 CGCAGCGTGA CCGCTACACT TGCCAGCGCC CTAGCGCCCG CTCCTTTCGC  
TTTCTTCCCT

3721 TCCTTTCTCG CCACGTTTCGC CGGCTTTCCC CGTCAAGCTC TAAATCGGGG  
GCTCCCTTTA

3781 GGGTTCCGAT TTAGTGATTT ACGGCACCTC GACCCCAAAA AACTTGATTA  
GGGTGATGGT

3841 TCACGTAGTG GGCCATCGCC CTGATAGACG GTTTTTCGCC CTTTGACGTT  
GGAGTCCACG

3901 TTCTTTAATA GTGGACTCTT GTTCCAAACT GGAACAACAC TCAACCCTAT  
CTCGGTCTAT

3961 TCTTTTGATT TATAAGGGAT TTTGCCGATT TCGGCCTATT GGTTAAAAAA  
TGAGCTGATT

4021 TAACAAAAAT TTAACGCGAA TTTTAACAAA ATATTAACGT TTACAATTTA  
AATATTTGCT

4081 TATACAATCT TCCTGTTTTT GGGGCTTTTC TGATTATCAA CCGGGGTACA  
TATGATTGAC

4141 ATGCTAGTTT TACGATTACC GTTCATCGAT TCTCTTGTTT GCTCCAGACT  
CTCAGGCAAT

4201 GACCTGATAG CCTTTGTAGA GACCTCTCAA AAATAGCTAC CCTCTCCGGC

ATGAATTTAT

4261 CAGCTAGAAC GGTGAATAT CATATTGATG GTGATTTGAC TGTCTCCGGC  
CTTTCTCACC

4321 CGTTTGAATC TTTACCTACA CATTACTCAG GCATTGCATT TAAAATATAT  
GAGGGTTCTA

4381 AAAATTTTTA TCCTTGCGTT GAAATAAAGG CTTCTCCCGC AAAAGTATTA  
CAGGGTCATA

4441 ATGTTTTTGG TACAACCGAT TTAGCTTTAT GCTCTGAGGC TTTATTGCTT  
AATTTTGCTA

4501 ATTCTTTGCC TTGCCTGTAT GATTTATTGG ATGTTGGAAG TTCCTGATGC  
GGTATTTTCT

4561 CCTTACGCAT CTGTGCGGTA TTTCACACCG CATATGGTGC ACTCTCAGTA  
CAATCTGCTC

4621 TGATGCCGCA TAGTTAAGCC AGCCCCGACA CCCGCCAACA CCCGCTGACG  
CGCCCTGACG

4681 GGCTTGTCTG CTCCCGGCAT CCGCTTACAG ACAAGCTGTG ACCGTCTCCG  
GGAGCTGCAT

4741 GTGTCAGAGG TTTTCACCGT CATCACCGAA ACGCGCGAGA CGAAAGGGCC  
TCGTGATACG

4801 CCTATTTTTA TAGGTTAATG TCATGATAAT AATGGTTTCT TAGACGTCAG  
GTGGCACTTT

4861 TCGGGGAAAT GTGCGCGGAA CCCCTATTG TTTATTTTTT TAAATACATT  
CAAATATGTA

4921 TCCGCTCATG AGACAATAAC CCTGATAAAT GCTTCAATAA TATTGAAAAA  
GGAAGAGTAT

4981 GAGTATTCAA CATTTCCGTG TCGCCCTTAT TCCCTTTTTT GCGGCATTTT  
GCCTTCCTGT

5041 TTTTGCTCAC CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT  
TGGGTGCACG

5101 AGTGGGTTAC ATCGAACTGG ATCTCAACAG CGGTAAGATC CTTGAGAGTT  
TTCGCCCCGA

5161 AGAACGTTTT CCAATGATGA GCACTTTTAA AGTTCTGCTA TGTGGCGCGG  
TATTATCCCG

5221 TATTGACGCC GGGCAAGAGC AACTCGGTCG CCGCATAAC TATTCTCAGA  
ATGACTTGGT

5281 TGAGTACTCA CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA  
GAGAATTATG

5341 CAGTGCTGCC ATAACCATGA GTGATAACAC TGCGGCCAAC TTAATTCTGA  
CAACGATCGG

5401 AGGACCGAAG GAGCTAACCG CTTTTTTGCA CAACATGGGG GATCATGTAA  
CTCGCCTTGA

5461 TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT ACCAAACGAC GAGCGTGACA  
CCACGATGCC

5521 TGTAGCAATG GCAACAACGT TGCGCAAACT ATTAAGTGGC GAACTACTTA  
CTCTAGCTTC

5581 CCGGCAACAA TTAATAGACT GGATGGAGGC GGATAAAGTT GCAGGACCAC  
TTCTGCGCTC

5641 GGCCCTTCCG GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC  
GTGGGTCTCG

5701 CGGTATCATT GCAGCACTGG GGCCAGATGG TAAGCCCTCC CGTATCGTAG  
TTATCTACAC

5761 GACGGGGAGT CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA  
TAGGTGCCTC

5821 ACTGATTAAG CATTGGTAAC TGTCAGACCA AGTTTACTCA TATATACTTT  
AGATTGATTT

5881 AAAACTTCAT TTTTAATTTA AAAGGATCTA GGTGAAGATC CTTTTTGATA  
ATCTCATGAC

5941 CAAAATCCCT TAACGTGAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG  
AAAAGATCAA

6001 AGGATCTTCT TGAGATCCTT TTTTTCTGCG CGTAATCTGC TGCTTGCAAA  
CAAAAAAACC

6061 ACCGCTACCA GCGGTGGTTT GTTTGCCGGA TCAAGAGCTA CCAACTCTTT  
TTCCGAAGGT

6121 AACTGGCTTC AGCAGAGCGC AGATACCAA TACTGTCCTT CTAGTGTAGC  
CGTAGTTAGG

6181 CCACCACTTC AAGAACTCTG TAGCACC GCG TACATACCTC GCTCTGCTAA  
TCCTGTTACC

6241 AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGA CTCAA  
GACGATAGTT

6301 ACCGGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGGTTCG TGCACACAGC  
CCAGCTTGGA

6361 GCGAACGACC TACACCGAAC TGAGATACCT ACAGCGTGAG CTATGAGAAA  
GCGCCACGCT

6421 TCCCGAAGGG AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCGGAA  
CAGGAGAGCG

6481 CACGAGGGAG CTTCCAGGGG GAAACGCCTG GTATCTTTAT AGTCCTGTGC  
GGTTTCGCCA

6541 CCTCTGACTT GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC

TATGGAAAAA

6601 CGCCAGCAAC GCGGCCTTTT TACGGTTCCT GGCCTTTTGC TGGCCTTTTG  
CTCACATGTT

6661 CTTTCCTGCG TTATCCCCTG ATTCTGTGGA TAACCGTATT ACCGGGTTTG  
AGTGAGCTGA

6721 TACCGCTCGC CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGACC  
AAGCGGAAGA

6781 GC。

[0169] SEQ ID NO 4

MSRSGTDPQQRQASEADAAAATFRANDHQHIRYNPLQDEWVLSAHRMKRPWQQVPEPQLLKTVPRH  
DPLNPLCPGAIRANGEVNPQYDSTFLFDNDFPALQPDAPSPGSDHPLFQAQSARGVCKVMCFHPWSDVTLPLMSV  
PEIRAVVDAWASVTEELGAQYPWVQIFENKGAMMGCSNPHPHCQVWASSFLPDIAQREERSQQAYKSQHGEPLLE  
YSRQELLRKERLVLVSEHWLVLVPFWATWPYQTLTLLPRRHVRRLPELTPAERDDLASIMKLLTKYDNLFETSFPY  
SMGWHGAPTGSEAGANWNHWQLHAHYYPLLRSATVRKFMVGYEMLAQQRDLTPEQAAERLRLALPEVHYHLGQKD  
RETATIA。

[0170] SEQ ID NO 5:

Rh74 VP1氨基酸序列

MAADGYLPDWLEDNLSEGIREWDLKPGAPKPKANQQKQDNGRGLVLPGYKYLGPFNGLDKGEPVNAAD  
AAALEHDKAYDQQLQAGDNPYLRYNHADADEFQERLQEDTSFGGNLGRAVFQAKKRVLEPLGLVESPVKTAPGKKRPV  
EPSPQRSPDSSTGIGKKGQQPAKKRLNFGQTGDSSEVPDPQPIGEPPAGPSGLSGTMAAGGGAPMADNNEGADGVG  
SSSGNWHCDSTWLGDREVITSTRTWALPTYNNHLYKQISNGTSGGSTNDNTYFGYSTPWGYDFNRFHCHFSRDPWQ  
RLINNNWGFPRKRLNFKLFNIQVKEVTQNEGTKTIANNLTSTIQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMIPQ  
YGYLTL

NGSQAVGRSSFYCLEYFPSQMLRTGNNFEFSYNFEDVVPFHSSYAHSQSLDRLMNPLIDQYLYLSRT  
QSTGGTAGTQQLLFSQAGPNNMSAQAKNWLPGPCYRQQRVSTTLSQNNNSNFAWTGATKYHLNGRDSLVPNGVAMA  
THKDDEERFFPSSGVLMFGKQGAGKDNVDYSSVMLTSEEEIKTTNPVATEQYGVVADNLQQQNAAP1VGAVNSQGA  
LPGMVWQNRDVYLQGPWAK1PHTDGNFHPSPLMGGFGLKHPPPQILIKNTPVPADPPTTFNQAKLASFITQYSTG  
QVSVEIEWELQKENSQRWNPEIQYTSNYYKSTNVDFAVNTEGTYSEPRPIGTRYLTRNL。

[0171] SEQ ID NO: 6

Rh74 VP1 DNA

ATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAGGGCATTTCGCGAGTGGTG  
GGACCTGAAACCTGGAGCCCCGAAACCCAAAGCCAACCAGCAAAAGCAGGACAACGGCCGGGTCTGGTGCTTCTCT  
GGCTACAAGTACCTCGGACCCTTCAACGGACTCGACAAGGGGGAGCCCGTCAACGCGCGGACGCAGCGGCCCTCG  
AGCAGGACAAGGCTACGACCAGCAGCTCCAAGCGGGTGACAATCCGTACCTGCGGTATAATCACGCCGACGCCGA  
GTTTCAGGAGCGTCTGCAAGAAGATACGTCTTTTGGGGCAACCTCGGGCGCGCAGTCTTCCAGGCCAAAAAGCGG  
GTTCTCGAACCTCTGGGCTGGTTGAATCGCCGGTTAAGACGGCTCCTGGAAAAGAGACCGGTAGAGCCATCAC  
CCCAGCGCTCTCCAGACTCCTCTACGGGCATCGGCAAGAAAGGCCAGCAGCCCGCAAAAAAGAGACTCAATTTTGG  
GCAGACTGGCGACTCAGAGTCAGTCCCCGACCCTCAACCAATCGGAGAACCACCAGCAGGCCCTCTGGTCTGGGA  
TCTGGTACAATGGCTGCAGGCGGTGGCGCTCCAATGGCAGACAATAACGAAGGCGCCGACGGAGTGGGTAGTTCCT

CAGGAAATTGGCATTGCGATTCCACATGGCTGGGCGACAGAGTCATCACCACCAGCACCCGCACCTGGGCCCTGCC  
CACCTACAACAACCACCTCTACAAGCAAATCTCCAACGGGACCTCGGGAGGAAGCACCAACGACAACACCTACTTC  
GGCTACAGCACCCCTGGGGTATTTTACTTCAACAGATTCCACTGCCACTTTTCACCACGTGACTGGAGCGACT  
CATCAACAACAACCTGGGGATTCCGGCCCAAGAGGCTCAACTTCAAGCTCTTCAACATCCAAGTCAAGGAGGTACAG  
CAGAATGAAGGCACCAAGAGCATCGCCAATAACCTTACCAGCAGGATTCAGGTCTTTACGGACTCGGAATACCAGC  
TCCCGTACGTGCTCGGCTCGGCGCACCAGGGCTGCCTGCCTCCGTTCCCGGCGGACGTCTTCATGATTCCCTAGTA  
CGGGTACCTGACTCTGAACAATGGCAGTCAGGCTGTGGGCCGGTCGTCCTTCTACTGCCTGGAGTACTTTCTTCT  
CAAATGCTGAGAACGGGCAACAACCTTTGAATTCAGCTACAACCTTCGAGGACGTGCCCTTCCACAGCAGCTACGCGC  
ACAGCCAGAGCCTGGACCGGCTGATGAACCCTCTCATCGACCAGTACTTGTACTACCTGTCCCGGACTCAAAGCAC  
GGGCGGTACTGCAGGAACTCAGCAGTTGCTATTTTCTCAGGCCGGGCTAACAACATGTCGGCTCAGGCCAAGAAC  
TGGCTACCCGGTCCCTGCTACCGGCAGCACGCTCTCCACGACACTGTCGCAGAACAACAACAGCAACTTTGCCTG  
GACGGGTGCCACCAAGTATCATCTGAATGGCAGAGACTCTCTGGTGAATCCTGGCGTTGCCATGGCTACCCACAAG  
GACGACGAAGAGCGATTTTTTCCATCCAGCGGAGTCTTAATGTTTGGGAAACAGGGAGCTGGAAAAGACAACGTGG  
ACTATAGCAGCGTGATGCTAACCAGCGAGGAAGAAATAAAGACCACCAACCCAGTGGCCACAGAACAGTACGGCGT  
GGTGGCCGATAACCTGCAACAGCAAAACGCCGCTCCTATTGTAGGGGCCGTC AATAGTCAAGGAGCCTTACCTGGC  
ATGGTGTGGCAGAACCGGGACGTGTACCTGCAGGGTCCCATCTGGGCCAAGATTCCTCATAACGACGGCAACTTTC  
ATCCCTCGCCGCTGATGGGAGGCTTTGGACTGAAGCATCCGCCTCCTCAGATCCTGATTA AAAACACACCTGTTCC  
CGCGGATCCTCCGACCACCTTCAATCAGGCCAAGCTGGCTTCTTTCATCACGCAGTACAGTACCGGCCAGGTCAGC  
GTGGAGATCGAGTGGGAGCTGCAGAAGGAGAACAGCAAACGCTGGAACCCAGAGATTCAGTACACTTCCA ACTACT  
ACAAATCTACAAATGTGGACTTTGCTGTCAATACTGAGGGTACTTATTCCGAGCCTCGCCCCATTGGCACCCGTTA  
CCTCACCCGTAATCTGTAA。

[0172] SEQ ID NO: 7 - KAN基因翻译产物

MSHIQRETSRPRLSNMDADLYGKWARDNVGQSGATIYRLYGKPDAPFLKHKGSVANDVTDE  
MVRLNWLTEFMPLPTIKHFIRTPDDAWLLTTAIPGKTAFAQVLEEYPSGENIVDALAVFLRRLHSIPVCNCPFNSD  
RVFRLAQAQSRMNGLVDASDFDDERNGWPVEQVWKEMHKLLPFSPDSVVTHGDFSLDNLIFDEGKLIGCIDVGRV  
GIADRYQDLAILWNCLGEFSPSLQKRLFQKYGIDNPDMNKLQFHLMLDEFF。

[0173] AAV rh74 共有序列比对

AAVrh74

AAVrh74-N502I-衣壳

Rh74 YIG591 帽蛋白。

共有序列  
 1. AAVrh74  
 2. AAVrh74-N502I-衣壳  
 3. rh74 YIG591 帽蛋白

[0174]

	*****										
	MAADGYLPDWLEDNLSEGIREWDLKPGAPKPKANQQKQDNGRGLVLPGYKYLGPFGNLDKGEPVNAADA										
	10	20	30	40	50	60	70				
1	MAADGYLPDWLEDNLSEGIREWDLKPGAPKPKANQQKQDNGRGLVLPGYKYLGPFGNLDKGEPVNAADA										70
2	MAADGYLPDWLEDNLSEGIREWDLKPGAPKPKANQQKQDNGRGLVLPGYKYLGPFGNLDKGEPVNAADA										70
3	MAADGYLPDWLEDNLSEGIREWDLKPGAPKPKANQQKQDNGRGLVLPGYKYLGPFGNLDKGEPVNAADA										70
	*****										
	AALEHDKAYDQQLQAGDNPYLRYNHADADEFQERLQEDTSFGGNLGRAVFAQAKRVLEPLGLVESPVKTAP										
	80	90	100	110	120	130	140				
1	AALEHDKAYDQQLQAGDNPYLRYNHADADEFQERLQEDTSFGGNLGRAVFAQAKRVLEPLGLVESPVKTAP										140
2	AALEHDKAYDQQLQAGDNPYLRYNHADADEFQERLQEDTSFGGNLGRAVFAQAKRVLEPLGLVESPVKTAP										140
3	AALEHDKAYDQQLQAGDNPYLRYNHADADEFQERLQEDTSFGGNLGRAVFAQAKRVLEPLGLVESPVKTAP										140
	*****										
	GKKRPVESPQRSPTSSTGIGKKGQPAKKRLNFGQTGDSESVDPQPIGEPAPGSPGLGSGTMAAGGGA										
	150	160	170	180	190	200	210				
1	GKKRPVESPQRSPTSSTGIGKKGQPAKKRLNFGQTGDSESVDPQPIGEPAPGSPGLGSGTMAAGGGA										210
2	GKKRPVESPQRSPTSSTGIGKKGQPAKKRLNFGQTGDSESVDPQPIGEPAPGSPGLGSGTMAAGGGA										210
3	GKKRPVESPQRSPTSSTGIGKKGQPAKKRLNFGQTGDSESVDPQPIGEPAPGSPGLGSGTMAAGGGA										210
	*****										
	PMADNNEGADGVGSSSGNWHCDSTWLGDRVITSTRTWALPTYNNHLYKQISNGTSGGSTNDNTYFGYST										
	220	230	240	250	260	270	280				
1	PMADNNEGADGVGSSSGNWHCDSTWLGDRVITSTRTWALPTYNNHLYKQISNGTSGGSTNDNTYFGYST										280
2	PMADNNEGADGVGSSSGNWHCDSTWLGDRVITSTRTWALPTYNNHLYKQISNGTSGGSTNDNTYFGYST										280
3	PMADNNEGADGVGSSSGNWHCDSTWLGDRVITSTRTWALPTYNNHLYKQISNGTSGGSTNDNTYFGYST										280
	*****										
	PWGYDFNRFHCHFSPRDWQRLINNNWGRPKRLNFKLFNIQVKEVTQNEGTKTIANNLTSTIQVFTDSE										
	290	300	310	320	330	340	350				
1	PWGYDFNRFHCHFSPRDWQRLINNNWGRPKRLNFKLFNIQVKEVTQNEGTKTIANNLTSTIQVFTDSE										350
2	PWGYDFNRFHCHFSPRDWQRLINNNWGRPKRLNFKLFNIQVKEVTQNEGTKTIANNLTSTIQVFTDSE										350
3	PWGYDFNRFHCHFSPRDWQRLINNNWGRPKRLNFKLFNIQVKEVTQNEGTKTIANNLTSTIQVFTDSE										350

	***** YQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGSQAVGRSSFYCLEYFPSQMLRTGNNFEFSYNFED *****	
	360 370 380 390 400 410 420	
1	YQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGSQAVGRSSFYCLEYFPSQMLRTGNNFEFSYNFED	420
2	YQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGSQAVGRSSFYCLEYFPSQMLRTGNNFEFSYNFED	420
3	YQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGSQAVGRSSFYCLEYFPSQMLRTGNNFEFSYNFED	420
	***** VPFHSSYAHSQSLDRLMNPLIDQYLYLSRTQSTGGTAGTQQLLFSQAGPNNMSAQAKNWLPGPCYRQQR *****	
	430 440 450 460 470 480 490	
1	VPFHSSYAHSQSLDRLMNPLIDQYLYLSRTQSTGGTAGTQQLLFSQAGPNNMSAQAKNWLPGPCYRQQR	490
2	VPFHSSYAHSQSLDRLMNPLIDQYLYLSRTQSTGGTAGTQQLLFSQAGPNNMSAQAKNWLPGPCYRQQR	490
3	VPFHSSYAHSQSLDRLMNPLIDQYLYLSRTQSTGGTAGTQQLLFSQAGPNNMSAQAKNWLPGPCYRQQR	490
	***** VSTTLSQNNNSNFAWTGATKYHLNGRDSLVPVAMATHKDDEERFFPSSGVL MFGKQGAGKDNVDYSSV *****	
	500 510 520 530 540 550 560	
1	VSTTLSQNNNSNFAWTGATKYHLNGRDSLVPVAMATHKDDEERFFPSSGVL MFGKQGAGKDNVDYSSV	560
2	VSTTLSQNNNSNFAWTGATKYHLNGRDSLVPVAMATHKDDEERFFPSSGVL MFGKQGAGKDNVDYSSV	560
3	VSTTLSQNNNSNFAWTGATKYHLNGRDSLVPVAMATHKDDEERFFPSSGVL MFGKQGAGKDNVDYSSV	560
	***** MLTSEEEIKTTNPVATEQYGVVADNLQQQNAAPIVGAVNSQGALPGMVWQNRDVYLQGGPIWAKIPHTDGN *****	
	570 580 590 600 610 620 630	
1	MLTSEEEIKTTNPVATEQYGVVADNLQQQNAAPIVGAVNSQGALPGMVWQNRDVYLQGGPIWAKIPHTDGN	630
2	MLTSEEEIKTTNPVATEQYGVVADNLQQQNAAPIVGAVNSQGALPGMVWQNRDVYLQGGPIWAKIPHTDGN	630
3	MLTSEEEIKTTNPVATEQYGVVADNLQQQNYIGSRGAVNSQGALPGMVWQNRDVYLQGGPIWAKIPHTDGN	630
	***** FHPSPLMGGFGLKHPPPQILIKNTPVPADPPTTFNQAKLASFITQYSTGQVSVEIEWELQKENS KRWNPE *****	
	640 650 660 670 680 690 700	
1	FHPSPLMGGFGLKHPPPQILIKNTPVPADPPTTFNQAKLASFITQYSTGQVSVEIEWELQKENS KRWNPE	700
2	FHPSPLMGGFGLKHPPPQILIKNTPVPADPPTTFNQAKLASFITQYSTGQVSVEIEWELQKENS KRWNPE	700
3	FHPSPLMGGFGLKHPPPQILIKNTPVPADPPTTFNQAKLASFITQYSTGQVSVEIEWELQKENS KRWNPE	700
	***** IQYTSNYYKSTNVDFAVNTEGTYSEPRPIGTRYLTRNL *****	
	710 720 730	
1	IQYTSNYYKSTNVDFAVNTEGTYSEPRPIGTRYLTRNL	738
2	IQYTSNYYKSTNVDFAVNTEGTYSEPRPIGTRYLTRNL	738
3	IQYTSNYYKSTNVDFAVNTEGTYSEPRPIGTRYLTRNL	738

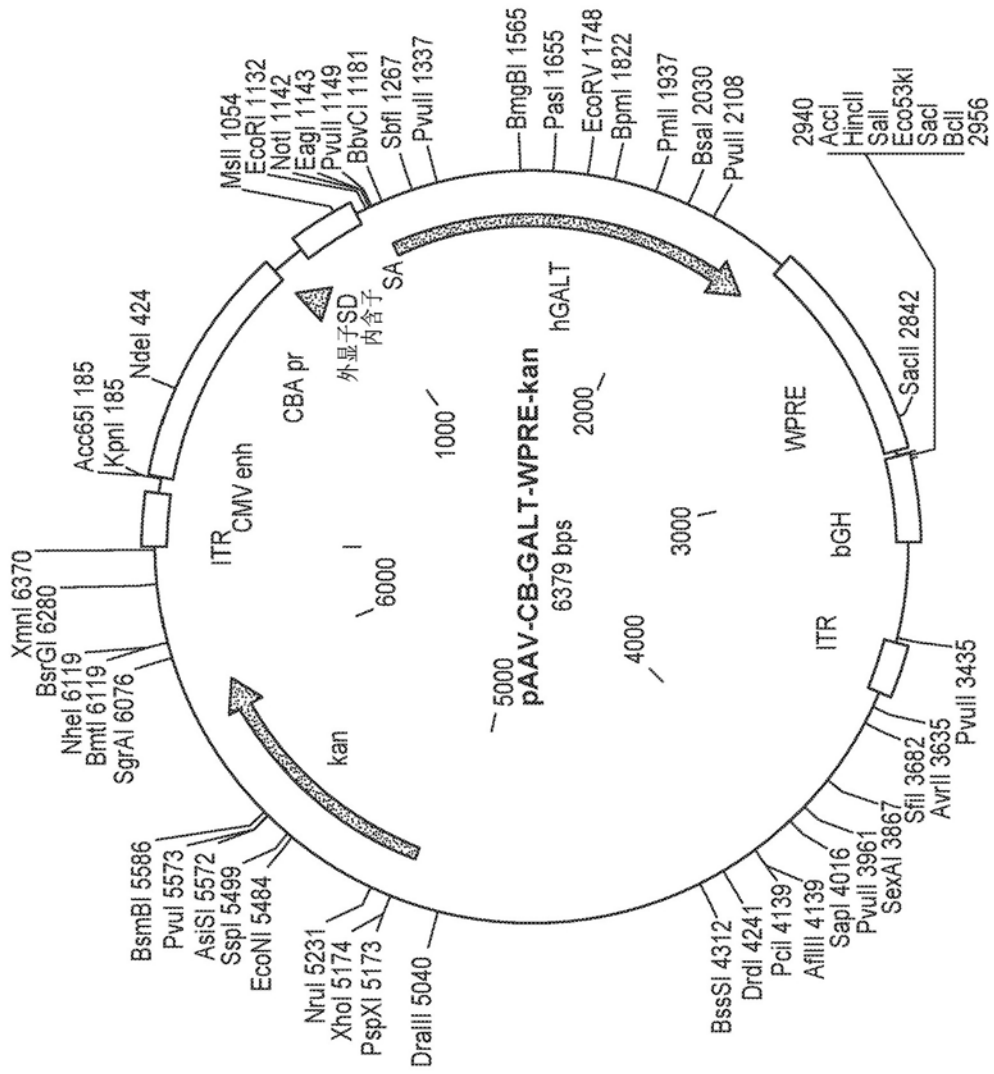


图1A

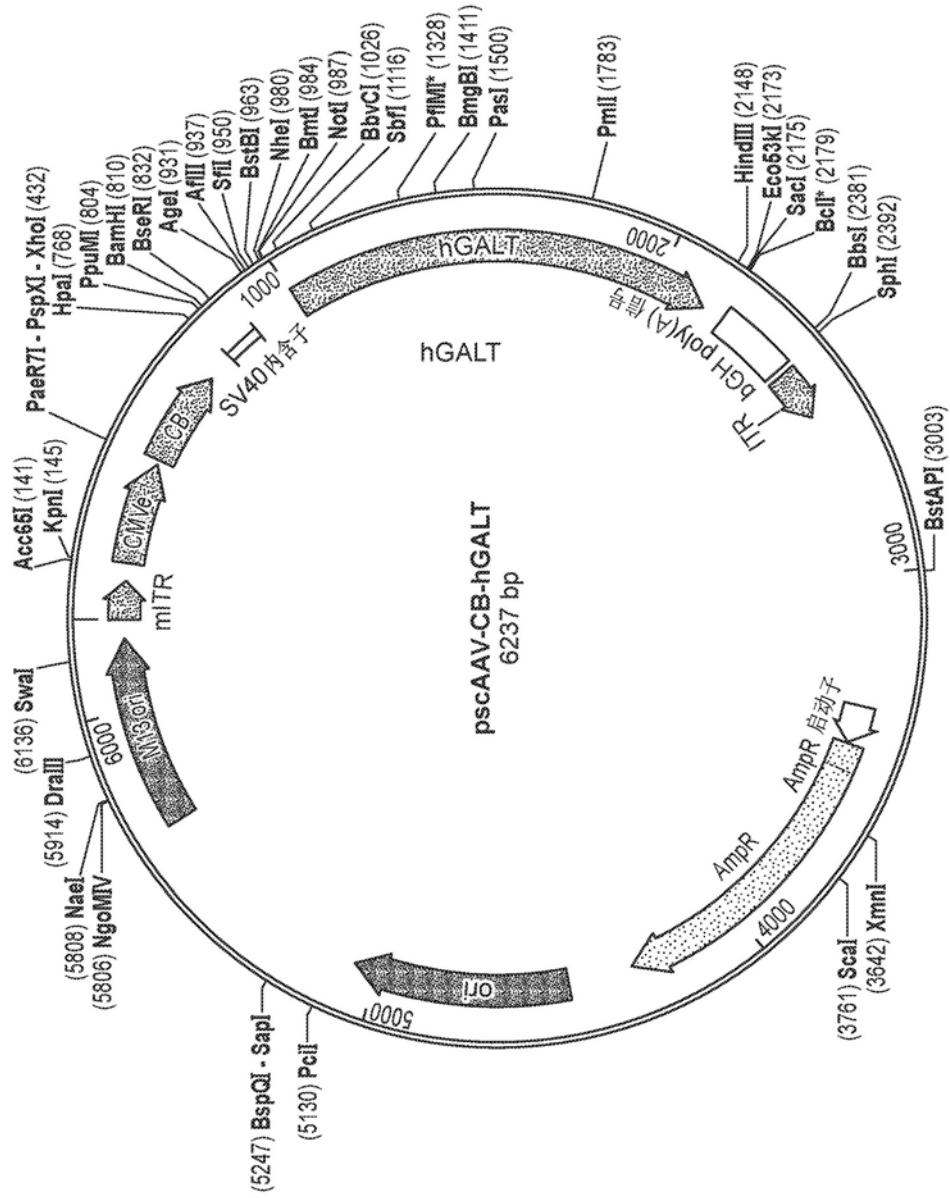


图1B



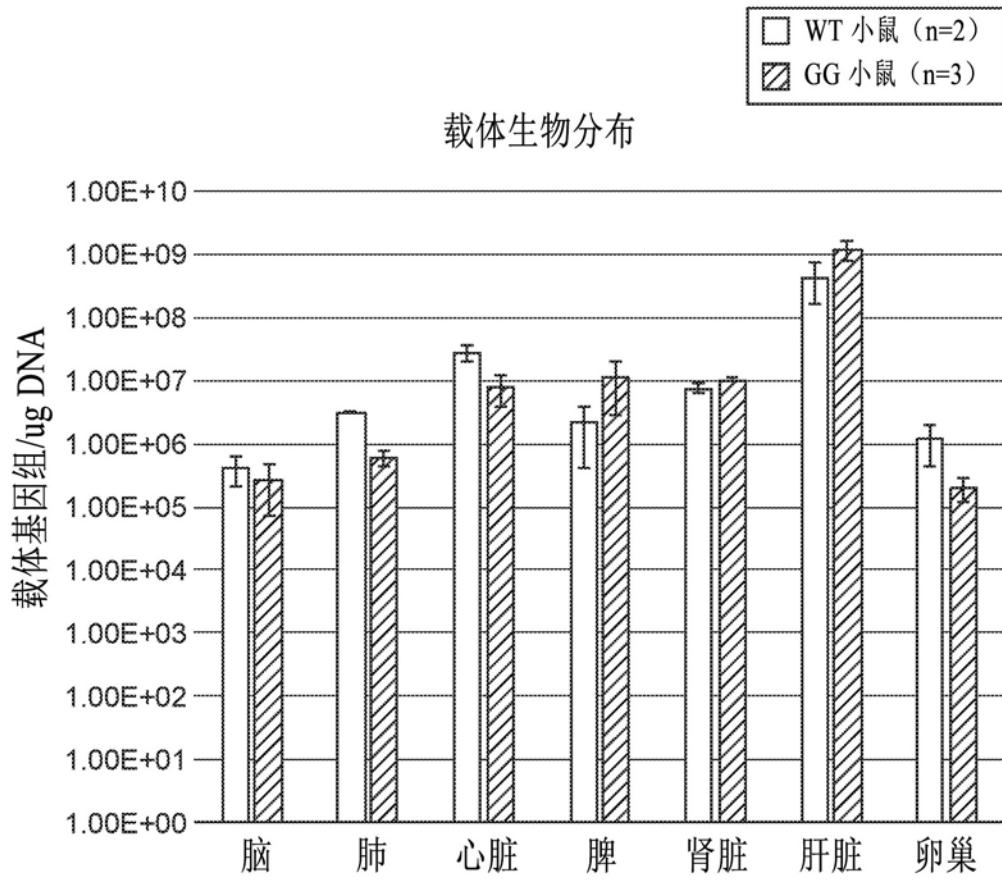


图2

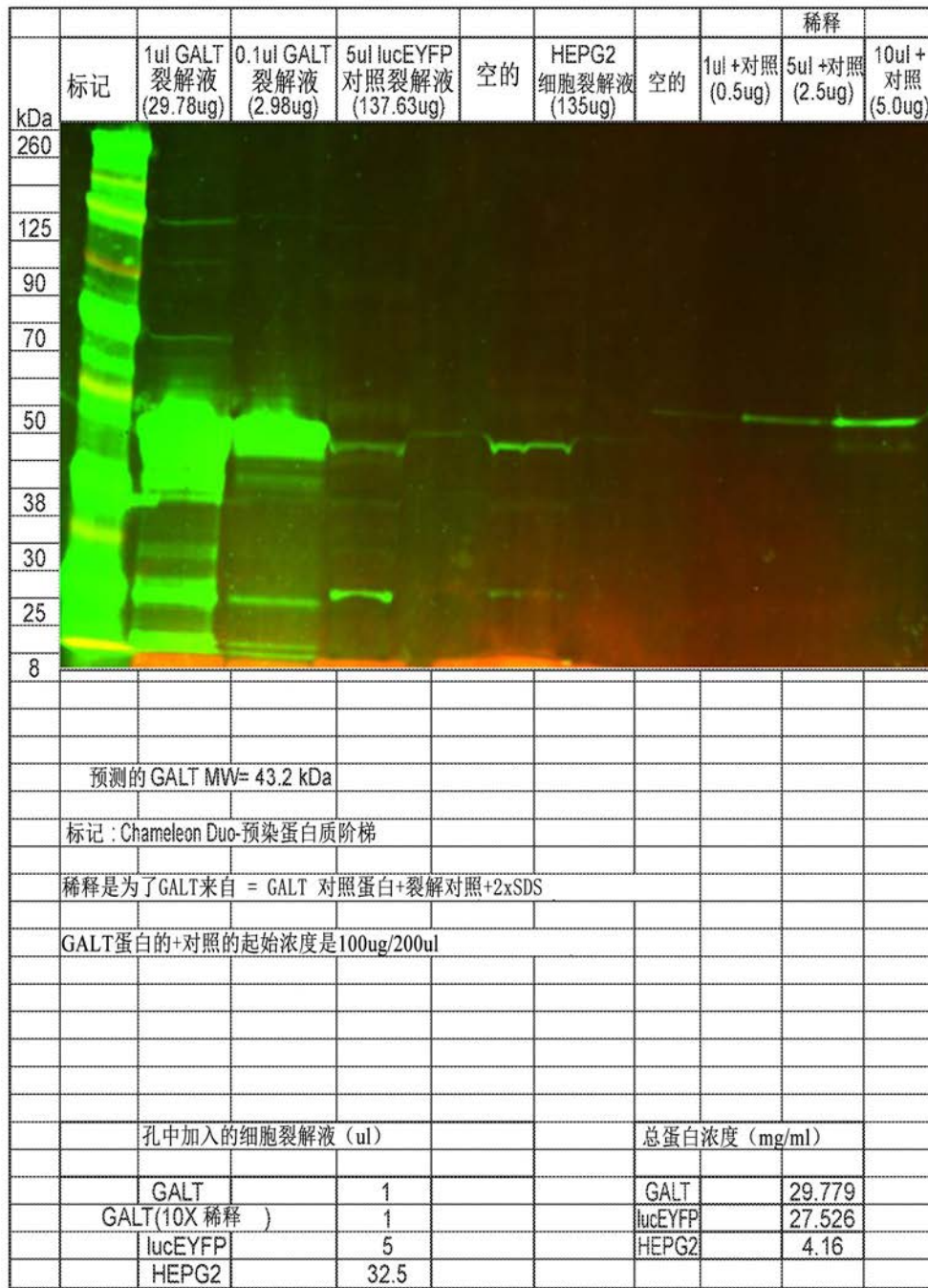


图3



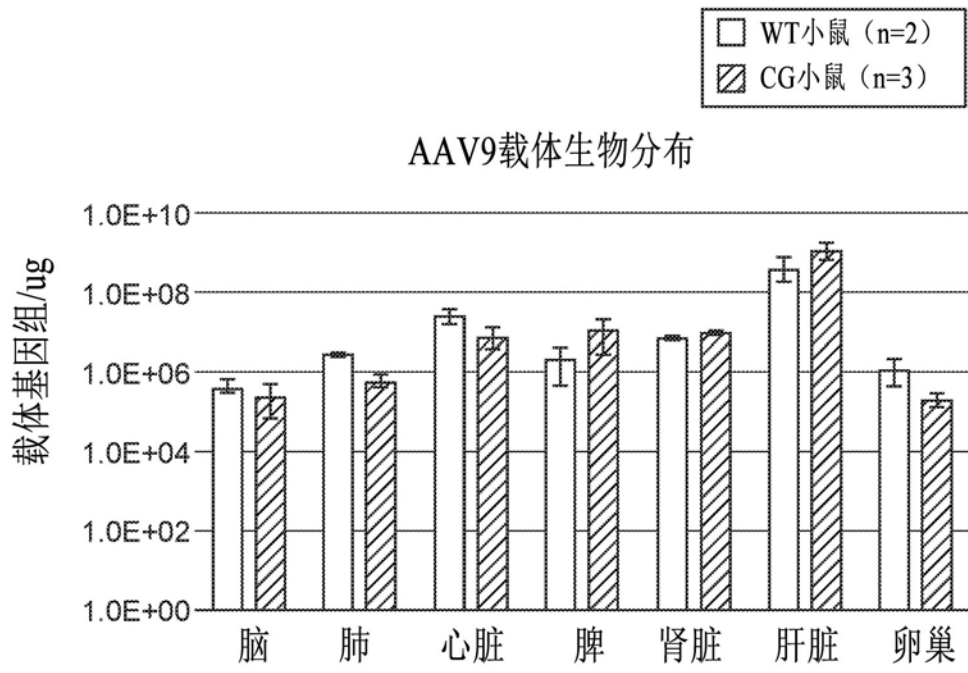


图5

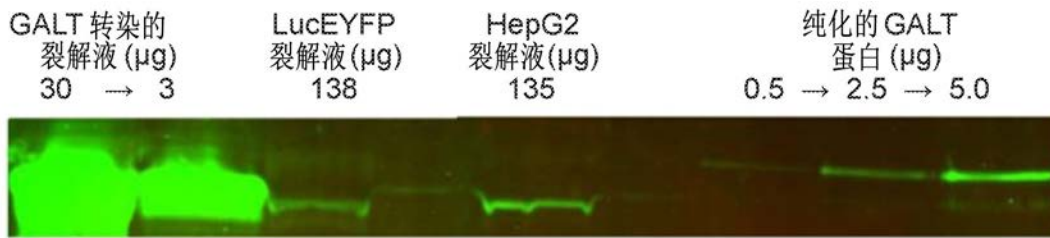


图6